

Article

Generative AI Literacy and Multidimensional AI Anxiety in Creative Higher Education

Xin Yong ^{1,2}, Junjie Gavin Wu ^{3,*} and Wei Wei ¹ 

¹ Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR, China

² School of Media, Sichuan Conservatory of Music, Chengdu 610500, China

³ Faculty of Languages and Translation, Macao Polytechnic University, Macao SAR, China

* Correspondence: gavinjunjiewu@gmail.com

Abstract

Engagement with generative artificial intelligence (GenAI) involves practical literacy as well as learning, professional, sociotechnical, and configuration concerns. This mixed-methods study examined whether these responses were differentiated among Chinese arts-university students. Primary inference used a quality-screened cross-sectional sample of 355 questionnaire records from 351 participant-identifier clusters after excluding 105 of 460 eligible records (22.8%) showing exact whole-battery straightlining across all 59 structured items; the full eligible sample of 460 records from 455 clusters was retained as a sensitivity analysis. GenAI literacy was assessed with a 16-item music-context-adapted self-report instrument covering cognitive (CL), behavioral (BL), affective (AL), and ethical (EL) literacy, and AI Anxiety with 21 items spanning Learning Anxiety (LA), Job-Replacement Anxiety (JR), Sociotechnical Blindness Anxiety (ST), and AI Configuration Anxiety (AC). Of six prespecified associations, only EL-JR and EL-ST met the Holm-adjusted criterion in the primary sample and remained supported in the full-sample and prior-use-adjusted sensitivity analyses. CL-LA and BL-LA were unsupported in both samples, whereas AL-LA and AL-AC were sample-specification sensitive. Measurement models showed mixed fit; CL-BL separation improved after quality screening but remained incomplete. A secondary 47-person exact-identifier matched panel showed higher later-wave CL, BL, and AL scores, whereas EL and all four AI Anxiety dimensions showed no Holm-adjusted difference. These literacy comparisons are exploratory because the administered literacy items were music-framed while the documented learning activities involved broadcasting scripts and AI-supported video production. Independently analyzed qualitative material provided construct-level consistency with the EL-ST interpretation and tentative complementarity for EL-JR, but was not participant-linked to the quantitative samples. Overall, the evidence supports a bounded interpretation of differentiated, calibrated rather than anxiety-free adaptation; it does not establish causal intervention effects or participant-level mixed-method mechanisms.

Keywords: generative artificial intelligence; AI literacy; AI anxiety; arts education; creative learning; mixed methods; psychological adaptation; educational psychology



Academic Editor: Olufemi A. Omitaomu

Received: 11 September 2026

Revised: 22 September 2026

Accepted: 26 September 2026

Published: 11 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. GenAI as an Educational-Psychology Problem

Generative artificial intelligence (GenAI) has rapidly become part of routine higher-education work, with students using generative systems to draft text, produce images and

video, develop ideas, and support discipline-specific problem solving. The educational significance of this shift extends beyond whether students adopt the technology. It concerns how learners understand GenAI, how they regulate their interaction with it, and how changes in perceived competence coexist with motivational, emotional, ethical, and professional responses. Mechanism-oriented research in educational psychology illustrates why this distinction matters. For example, responsive conversational AI can improve learning when it changes the quality of interaction and feedback rather than simply adding a technological tool [1]. More broadly, GenAI-supported writing, digital storytelling, music, visual art, and design studies show that generative systems can alter both learning processes and students' judgments about creativity, authorship, and human contribution [2–6]. Accordingly, the present study treats GenAI as an educational-psychology problem involving cognition, behavior, affect, and appraisal rather than as a technology-acceptance problem alone.

This framing is especially important because psychologically meaningful outcomes do not necessarily move in the same direction. A learner may become more capable of using GenAI while remaining concerned about system errors, professional displacement, or the broader consequences of automation. Conversely, anxiety can sometimes accompany greater vigilance or motivated engagement rather than simple avoidance. Recent studies have reported negative associations between AI literacy and teaching anxiety, positive relations between some forms of AI anxiety and motivated learning, and the coexistence of high literacy with high anxiety [7–10]. These patterns make it difficult to justify a simple educational assumption that greater literacy should uniformly produce lower anxiety. They instead motivate a dimension-sensitive account of how capability and concern may be organized together.

1.2. Multidimensional GenAI Literacy: From Exposure to Calibrated Competence

AI literacy is increasingly conceptualized as a multidimensional educational competence rather than as frequency of use or familiarity with a particular application. The 17-item GenAI literacy scale developed for digital multimodal composing by Liu et al. [11], for example, distinguishes cognitive, behavioral, affective, and ethical dimensions. Complementary frameworks also emphasize that literacy includes knowledge, higher-order thinking, affective orientations, and the human and societal implications of AI. Zhong and Liu [12] organized secondary students' AI literacy around AI Knowledge, AI Affectivity, and AI Thinking, with the affective component explicitly including AI–human and AI–society relations. Such work supports the broader proposition that learners can be relatively strong in one aspect of AI literacy while remaining less developed in another.

A second distinction concerns perceived literacy versus demonstrated competence. Reviews of AI-literacy instruments show that self-report measures are much more common than performance-based assessments and that evidence for cross-cultural validity, responsiveness, and measurement error remains limited across many instruments [13]. Performance-based work has therefore been important for separating what learners believe they know from what they can demonstrate. Hornberger et al. [14], for instance, developed a multiple-choice AI literacy test for university students rather than relying solely on self-assessment. The present study does not claim such performance measurement. Its music-context-adapted 16-item instrument is explicitly treated as a self-report measure of cognitive (CL), behavioral (BL), affective (AL), and ethical (EL) GenAI literacy, and its local measurement properties are evaluated rather than assumed from the closest source scale.

Recent GenAI research further suggests that exposure and use should not be equated with calibrated understanding. Jaggy et al. [15] showed that a notable minority of secondary students retained an “omniscient-AI” misconception and that students differed in their

likelihood of revising inaccurate conceptions. Although that study did not measure AI anxiety, it provides an educational-psychology basis for interpreting cognitive literacy as more than technical familiarity: calibrated literacy involves understanding both capabilities and limitations. Likewise, a large randomized study of 1368 university students found that short GenAI learning modules improved knowledge, prompt-engineering skill, fact- and source-checking skill, and self-efficacy, while critical evaluation of potential bias did not show the same improvement [16]. Competence development can therefore be uneven across operational and critical dimensions.

These distinctions inform several of the study's prespecified associations. CL was paired with Learning Anxiety (LA) because more calibrated understanding could plausibly relate to uncertainty about learning and keeping pace with AI, but existing evidence does not warrant assuming the direction of that relation. BL was also paired with LA because repeated use and operational engagement may reduce task friction while remaining psychologically separable from broader anxiety. EL was paired with Job-Replacement Anxiety (JR) and Sociotechnical Blindness Anxiety (ST) because ethical literacy directly engages questions of human agency, misuse, bias, misinformation, responsibility, and societal consequences. These pairings are theory-guided construct links, not claims that one dimension causes another.

1.3. AI Anxiety as a Heterogeneous and Dynamically Embedded Response

AI anxiety is similarly better treated as multidimensional than as a single negative state. Wang and Wang's [17] Artificial Intelligence Anxiety Scale (AIAS) distinguishes Learning Anxiety, Job-Replacement Anxiety, Sociotechnical Blindness Anxiety, and Configuration Anxiety. LA concerns learning, understanding, operating, and keeping pace with AI; JR concerns dependency, reduced reasoning, and human or job replacement; ST concerns misuse, malfunction, loss of control, and autonomous systems; and Configuration Anxiety (AC) concerns fear or discomfort elicited by human-like AI technologies or products, such as humanoid robots. The administered AC items do not measure software configuration, prompting, or output-control difficulty. Research using this multidimensional tradition increasingly indicates that the dimensions need not have identical correlates or functions. Network evidence from Chinese users, for example, found that AI literacy and information overload related differently to distinct anxiety dimensions, with ST emerging as a particularly central node and LA as an important bridge [18]. Teke [9] likewise identified different AI-anxiety profiles among preservice teachers and reported that a high risk-awareness profile could coexist with stronger motivated learning.

The broader educational-psychology literature provides a useful mechanism-level caution against treating anxiety as uniformly maladaptive. In math learning, the consequences of anxiety for attention and performance vary with emotion regulation and state motivation [19]. Longitudinal work on test anxiety shows reciprocal relations between anxiety and achievement, indicating that temporal direction can run both ways rather than from a presumed antecedent to a fixed outcome [20]. Reading research similarly situates anxiety within a larger system of self-efficacy, cognitive processes, engagement, and achievement [21]. These studies are not AI-anxiety studies and are not used here as direct evidence for the present paths. Their relevance is conceptual: educational anxiety is often embedded in appraisal, regulation, motivation, and engagement processes, so the meaning of an anxiety score cannot be inferred from valence alone.

This logic is particularly relevant to the two affective-literacy paths specified in the present study. AL was paired with LA because positive expectations, perceived ease, personalization, and creativity-related orientations may covary with how threatening or manageable AI learning feels. The AL-AC association examines whether favorable GenAI

appraisals covary with fear of human-like AI technologies or products. This interpretation follows the administered item content; theoretical support for this specific pairing is indirect, and no directional relation is assumed. Both AL-LA and AL-AC were specified as two-sided associations, preserving the possibility of null, positive, negative, or unstable patterns.

1.4. Literacy and Anxiety: Beyond a Uniform Buffering Hypothesis

Research directly connecting AI literacy and AI anxiety now makes a simple “more literacy, less anxiety” claim difficult to defend. In university students, Cengiz and Peker [22] found that AI literacy formed part of an acceptance-related pathway associated with lower AI anxiety. Qiu et al. [18] similarly identified AI literacy as an anxiety-mitigating factor within a network of media exposure and information overload. In teacher samples, higher AI literacy has been associated with lower teaching anxiety through self-efficacy and engagement processes [7]. These findings are consistent with an uncertainty-reduction mechanism in which knowledge and perceived competence increase control and reduce some forms of apprehension.

However, other studies demonstrate that this buffering account is incomplete. Zhuge and Li [10] reported a positive association between AI anxiety and AI literacy and used interviews with high-anxiety/high-literacy students to interpret anxiety partly as vigilance-driven engagement. Xu et al. [23], using two-wave data from 408 digital-art undergraduates, showed that AI literacy had functionally differentiated relations with learning, sociotechnical, and job-replacement anxieties in motivational processes. In that study, reducing some anxieties supported intrinsic motivation, whereas reducing job-replacement anxiety could also weaken an extrinsic, survival-oriented motivational pathway. These results do not imply that anxiety is desirable; they indicate that different forms of anxiety can carry different informational and motivational meanings.

Intervention and experimental evidence reinforces the need to decouple competence growth from anxiety reduction. Hsu et al. [24] found that an experiential voice-assistant learning approach improved computational thinking, but the groups did not differ significantly in AI concepts or AI anxiety. Connell Pensky et al. [16] found clear gains in several GenAI competency components but not in critical evaluation of bias. In creative-thinking research, GenAI assistance improved creative performance while increasing fear-related emotion in a brainstorming condition [25]. Fear in that study is not equivalent to the validated AI Anxiety dimensions used here, but the coexistence of better performance and negative affect provides a useful boundary example: capability gain and emotional reassurance are not interchangeable outcomes.

Taken together, the literature supports two nonexclusive mechanisms. An uncertainty-reduction mechanism predicts that knowledge, practice, and control can reduce some task- or learning-related anxieties. A critical-awareness mechanism predicts that greater engagement with AI can also make limitations, ethical issues, occupational disruption, and sociotechnical consequences more salient. The present study therefore does not posit a uniform buffering effect. Instead, it tests selected dimension-to-dimension associations and interprets their direction empirically. CL-LA and BL-LA speak most directly to the uncertainty-reduction possibility; AL-LA and AL-AC concern affective appraisal in relation to learning anxiety and fear of human-like AI products; and EL-JR and EL-ST address the possibility that critical or ethical engagement may coexist with professional and sociotechnical concern.

1.5. Arts and Creative Higher Education as a Psychologically Consequential Context

Arts and creative higher education provides a theoretically consequential context for examining these differentiated processes because GenAI acts directly on products that

students may regard as extensions of authorship, aesthetic judgment, expressive identity, and future professional value. Existing arts research already documents both benefit and concern. Students in digital storytelling have reported collaborative and creative advantages alongside concerns about overreliance, originality, and emotional depth [5]. Art-education students have described inspiration and increased experience with GenAI while retaining technical and originality concerns [2]. Studies of voice and visual-art contexts have similarly examined AI anxiety, expressive appraisal, creative identity, employment concern, authorship, and human–AI co-creation [26–28]. The arts population is therefore not novel by itself.

What makes the context theoretically useful is the proximity between GenAI capabilities and students’ developing professional identities and creative practices. Song et al. [25] showed experimentally that GenAI can enhance creative performance while producing different affective consequences across creative tasks. Jou and Hao [29] conceptualized AI substitution as a psychologically salient learning event and linked disruption-related anxiety, career awareness, skill-transformation acceptance, and self-directed learning under instructional scaffolding. Although their anxiety construct is not identical to the present JR dimension, the study is directly relevant to the educational meaning of perceived occupational disruption. In parallel, Zhong and Liu’s [12] inclusion of AI–human and AI–society relations within AI literacy underscores that societal and human consequences can be integral to how learners understand AI rather than external add-ons.

This context provides the strongest theoretical rationale for the EL-JR and EL-ST pairings. Students who attend closely to copyright, bias, misinformation, human agency, and responsible AI use may also attend more closely to the consequences of AI for professional roles and sociotechnical systems. That coexistence could reflect risk-oriented appraisal, reverse direction, shared exposure, semantic adjacency, or third variables; cross-sectional evidence cannot distinguish among these explanations. EL-JR is therefore interpreted as a potential link between ethical/professional awareness and replacement-related concern, while EL-ST captures the even more direct overlap between ethical attention and concern about misuse, malfunction, control, and societal consequences. Neither path is framed as evidence that ethical literacy creates anxiety.

1.6. Differentiated Psychological Adaptation and Calibrated AI Engagement

The preceding literature motivates differentiated psychological adaptation as a bounded organizing lens for the present study. The term does not denote a validated latent construct, a new general theory, or a causal mechanism. It summarizes a more limited proposition: successful adaptation to GenAI may involve growth in some forms of competence, selective relief of task burden, more reflective or calibrated use, and sustained ethical or professional concern at the same time. Under this lens, lower anxiety is not automatically evidence of better adaptation, and persistent concern is not automatically evidence of maladaptation.

Several recent studies make this proposition educationally plausible. Jaggy et al. [15] demonstrate why calibrated understanding of GenAI limitations matters. Wang et al. [19] show, in a non-AI domain, that the consequences of anxiety depend on regulatory and motivational conditions. Jou and Hao [29] illustrate how anxiety associated with technological disruption can function as a conditional signal within a scaffolded learning process. Song et al. [25] show that improved creative performance can coexist with fear-related affect. Hou et al. [30] further distinguish reflective, cautious, thoughtless, and collaborative forms of reliance on GenAI during problem solving, emphasizing that effective human–AI engagement concerns how learners rely on AI, not simply how much they use it. This

distinction aligns with process-oriented GenAI literacy work showing movement toward more selective and reflective use over time [6].

The lens also provides a principled bridge to the qualitative strand of the present study. Accounts of calibrated output use, human-led creative collaboration, and ethical and professional vigilance can contextualize quantitative patterns without being treated as substitutes for validated scale dimensions. Likewise, immediate production-task anxiety can be informative without being equated with LA, JR, ST, or AC. Differentiated adaptation therefore requires construct boundaries as well as multidimensionality: practical competence, general AI Anxiety, and task-specific emotion may coexist, but they should not be collapsed into a single continuum of “positive” versus “negative” adaptation.

1.7. Research Gap, Prespecified Path Family, and Study Contribution

The 2023–2026 literature substantially narrows the gap that can now be defended. It is no longer accurate to claim that few studies connect AI literacy and AI anxiety, that arts students have rarely been studied, or that mixed methods are novel in this area. Aggregate literacy-anxiety models, acceptance and mediation models, multidimensional AI-anxiety analyses, arts and design studies, and mixed-method investigations are all present in the recent literature [10,18,22,23]. The more defensible gap concerns how distinct literacy dimensions correspond to qualitatively different forms of AI anxiety and how those relations should be interpreted when response-quality sensitivity, null findings, bounded repeated-measures evidence, and construct-level qualitative integration are retained.

Four limitations in the existing literature are especially relevant. First, multidimensional constructs are often measured but subsequently summarized through total scores or broad mediation paths, which can obscure dimension-specific associations. Second, buffering or acceptance framings do not fully distinguish uncertainty reduction from critical awareness. Third, mixed-method studies can add explanatory depth only when construct and participant-linkage boundaries are explicit. Fourth, single-group repeated-measures evidence requires separation from causal intervention claims, particularly when measurement context and documented learning activities are not fully aligned.

Against this background, the D1 analysis specified six two-sided associations before substantive outcome testing: CL-LA, BL-LA, AL-LA, AL-AC, EL-JR, and EL-ST. The family was intentionally selective rather than a post hoc complete 4×4 model. CL-LA and BL-LA test whether cognitive calibration and behavioral engagement are associated with learning-related concern without presuming that competence must reduce it. AL-LA and AL-AC test whether affective orientation covaries with learning anxiety and fear of human-like AI products. EL-JR and EL-ST test whether ethical/professional attention covaries with replacement/dependency and sociotechnical concern. All six associations are interpreted as cross-sectional relations, not effects, and multiplicity is controlled across the family.

The study adds two secondary evidence layers without changing the inferential priority of D1. First, 47 exact identifier (ID) matches between datasets D4 and D5 provide ordered repeated-measures evidence across CL, BL, AL, EL, LA, JR, ST, and AC. The paired analysis is descriptive and secondary; the exact calendar chronology is unresolved, and the music-framed literacy items are not fully aligned with the documented broadcasting/video learning activities. Second, independently analyzed reflections and interviews are integrated at the construct, theme, cohort, or study-context level. They are not used to explain individual quantitative scores or trajectories where identifiers do not link the strands.

The study addresses three research questions:

RQ1. How are distinct dimensions of self-reported Generative AI literacy associated with different dimensions of AI Anxiety in the quality-screened D1 sample, and

which prespecified associations remain supported in the full-sample and covariate-adjusted sensitivity analyses?

RQ2. What within-person differences in the four GenAI literacy dimensions and four AI Anxiety dimensions are observed across the two ordered D4–D5 measurement waves in the 47 exact-ID matched panel?

RQ3. At the construct, theme, and study-context levels, how do independently analyzed qualitative accounts converge with, complement, diverge from, or remain silent about the quantitative patterns without implying participant-level correspondence?

2. Materials and Methods

2.1. Research Design

This study used a layered mixed-methods design to examine dimension-specific relations between self-reported GenAI literacy and AI Anxiety. The primary quantitative component was a cross-sectional analysis of D1. Inference was anchored in a quality-screened sample of 355 records from 351 participant-identifier clusters, with the full eligible sample of 460 records from 455 clusters retained as a response-quality sensitivity analysis. The six primary associations and their six-test Holm multiplicity family were specified in the retained internal analysis plan on 31 August 2026, before Phase 5 outcome testing. The project audit trail records the Phase 5 plan freeze at 06:31:47 Coordinated Universal Time (UTC) (14:31:47 Asia/Shanghai) and the response-quality sensitivity-rule freeze at 14:34:45 Asia/Shanghai. The subsequent Phase 5 quantitative analysis record states that outcome testing began after both freezes. These are internal project records, not an external preregistration; we therefore describe the six associations as internally prespecified primary associations.

The secondary quantitative component used 47 exact D4–D5 identifier matches to describe within-person differences across two ordered waves surrounding researcher-documented AI-supported creative learning. It was not designed to estimate an intervention effect. The qualitative component comprised reflection journals and semi-structured interview material analyzed independently of the quantitative results.

Integration occurred only after the strand-specific analyses had been completed. Because D9, D11, and Q1 were not verified as participant-linked to D1 or the D4–D5 matched panel, cross-strand integration was conducted at the construct, theme, cohort, and study-context levels rather than the participant level. Figure 1 summarizes the revised evidence hierarchy and linkage boundaries.

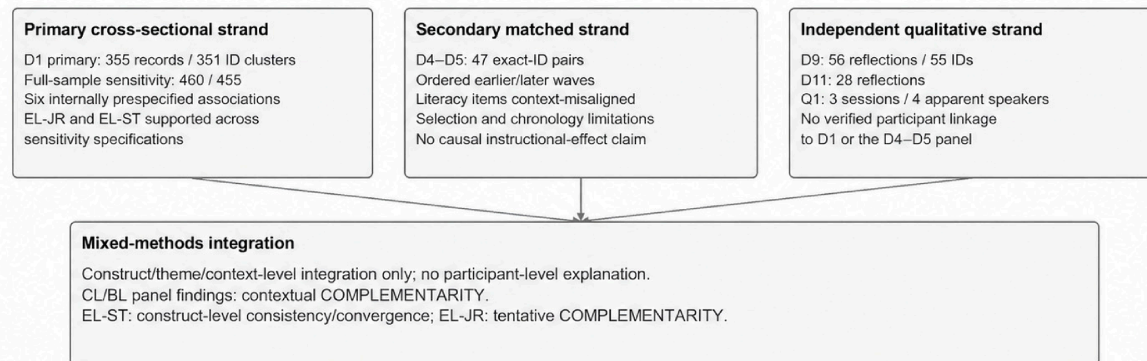
2.2. Participants and Educational Context

The project drew on questionnaire and qualitative records from undergraduate students aged 18 years or above at a Chinese arts university; no minors participated. The available records covered music and music education, music-related production tasks, broadcasting and television directing, broadcasting and hosting, and related creative-production contexts. The datasets were not treated as one fully linked cohort.

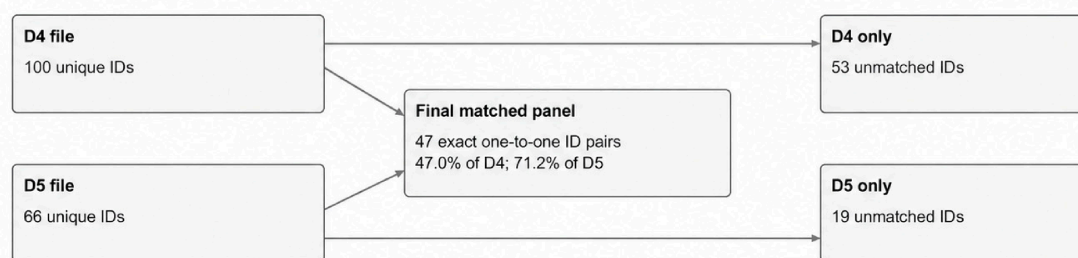
D1 contained 460 questionnaire records representing 455 participant identifiers. The exact-straightline response-quality rule described in Section 2.7 yielded the primary inferential sample of 355 records from 351 identifier clusters. The full eligible D1 sample was retained as a sensitivity analysis. Five identifiers occurred twice; all records were retained because the paired submissions were not identical and the available records did not identify either submission as invalid. Participant-identifier-clustered uncertainty estimates accounted for within-identifier dependence.

Study evidence and exact-ID matching

(A) Evidence strands and linkage boundary



(B) Exact identifier matching



Unmatched IDs indicate no exact cross-wave match, not verified dropout.

Figure 1. Revised study design, evidence architecture, and D4–D5 exact-ID matching flow. **(A)** The primary D1 strand, secondary matched strand, independent qualitative strand, and construct/theme/context-level integration boundary. **(B)** D4 = 100 unique IDs and D5 = 66 unique IDs yielding 47 exact one-to-one matches, 53 D4-only IDs, and 19 D5-only IDs. Unmatched identifiers denote absence of an exact cross-wave match and are not assumed to be verified dropouts. D1 denotes the primary questionnaire dataset; D4 and D5 denote the earlier and later questionnaire datasets; D9 and D11 denote reflection datasets; Q1 denotes the interview dataset. ID = identifier; AI = artificial intelligence; CL = cognitive literacy; BL = behavioral literacy; EL = ethical literacy; JR = job-replacement anxiety; ST = sociotechnical blindness anxiety.

The original D1 questionnaire collected participant identifier, major, gender, year of study, prior GenAI use for video production, and prior GenAI use for music production. Major was entered as free text. Gender, year, major, and both prior-use variables had no missing entries in the D1 export. No age or age-category item was present in the original D1 questionnaire, so age is not reported as a D1 demographic variable.

Among the 355 primary-sample records, 296 (83.4%) indicated female gender and 59 (16.6%) male gender. First-year records comprised 47.9%, second-year 36.3%, third-year 3.9%, and fourth-year 11.8%. The four most frequent raw major labels were Music Education (35.5%), Musicology (33.0%), Music Performance (12.4%), and Dance Performance (6.2%). Because the major field contained multiple abbreviated, discipline-level, and department-level free-text variants, no post hoc broad recoding of major was imposed for inferential analysis; the complete raw-label distribution is reported in Supplementary Table S2. Table 1 reports primary and full-sample participant characteristics.

Table 1. Participant characteristics of the primary quality-screened and full eligible D1 samples.

Full <i>n</i> = 460	Primary <i>n</i> = 355	Characteristic
370 (80.4%)	296 (83.4%)	Gender—Female
90 (19.6%)	59 (16.6%)	Gender—Male
202 (43.9%)	170 (47.9%)	Year 1
181 (39.3%)	129 (36.3%)	Year 2
15 (3.3%)	14 (3.9%)	Year 3
62 (13.5%)	42 (11.8%)	Year 4
155 (33.7%)	126 (35.5%)	Major: Music Education
163 (35.4%)	117 (33.0%)	Major: Musicology
53 (11.5%)	44 (12.4%)	Major: Music Performance
30 (6.5%)	22 (6.2%)	Major: Dance Performance
59 (12.8%)	46 (13.0%)	Major: Other free-text labels
74 (16.1%)	59 (16.6%)	Prior video-AI use: none
310 (67.4%)	234 (65.9%)	Prior video-AI use: occasional (1–2/month)
59 (12.8%)	50 (14.1%)	Prior video-AI use: frequent (1–4/week)
17 (3.7%)	12 (3.4%)	Prior video-AI use: very frequent (≥5/week)
221 (48.0%)	186 (52.4%)	Prior music-AI use: none
203 (44.1%)	141 (39.7%)	Prior music-AI use: occasional (1–2/month)
25 (5.4%)	22 (6.2%)	Prior music-AI use: frequent (1–4/week)
11 (2.4%)	6 (1.7%)	Prior music-AI use: very frequent (≥5/week)

Note. Percentages are within sample. Major was collected as free text; the four most frequent raw labels are shown and all remaining labels are grouped descriptively as Other. No post hoc broad major recoding was used in inferential analysis; the complete raw-label distribution is in Supplementary Table S2. No D1 age item was present. Variables shown had no missing entries. The primary sample contains 355 records from 351 participant-identifier clusters; the full sample contains 460 records from 455 clusters.

Prior GenAI experience was measured separately for video and music production using four categories: no prior use, occasional use (1–2/month), frequent use (1–4/week), and very frequent use (≥5/week). In the primary sample, the corresponding percentages were 16.6%, 65.9%, 14.1%, and 3.4% for video-related use, and 52.4%, 39.7%, 6.2%, and 1.7% for music-related use. Because prior exposure could plausibly be associated with both self-reported literacy and AI Anxiety, the two prior-use variables were additionally entered as categorical covariates in sensitivity analyses. Gender and year were reported descriptively but were not added post hoc to the prespecified primary system.

D4 contained 100 unique identifiers and D5 contained 66. Forty-seven identifiers matched exactly across the two files, with no duplicate or ambiguous matched pairs. Researcher documentation identified these 47 students as members of two intact classes that received the same instructional sequence. The exact official program/class identity could not be reconciled across all administrative/exported labels and context-specific item wording, so the manuscript does not assign the paired panel a single official discipline label.

For D1, D4, D5, D9, D11, and Q1, recruitment notices were posted in relevant course communication groups and participation was voluntary. Participation, refusal, or withdrawal did not affect grades, assessment, academic standing, or access to routine instruction,

and no financial, academic, or material compensation was provided. Eligibility required undergraduate status, age 18 years or above, enrollment in the applicable target course, and completion of the applicable informed-consent procedure.

2.3. AI-Supported Creative Learning Context

Researcher documentation described the D4–D5 instructional sequence as four approximately 30-min discipline-embedded activities, for a total of about 120 min. The first introduced GenAI and possible applications in Broadcasting and Hosting; the second demonstrated Doubao for broadcasting-script generation and Jimeng for AI-supported video generation; the third involved hands-on script and video production; and the fourth used guided reflection to address copyright, bias, ethics, professional implications, human creativity, and human agency.

A substantive context mismatch was identified during the data audit. The 16 GenAI literacy items administered at D4 and D5 were framed around music production, music learning, or music education, whereas the documented instructional activities centered on broadcasting/hosting scripts and AI-supported video production. The available records did not provide a defensible basis for retrospectively relabeling the literacy items or reconstructing a broadcasting-specific literacy scale.

This mismatch is therefore treated as a contextual-validity limitation of the repeated-measures literacy evidence. The same literacy items were administered at both waves and can be compared descriptively within participants, but differences in those scores cannot be interpreted as direct measurement of literacy acquired specifically through the documented broadcasting/video activities. The AI Anxiety items were domain-general and are less directly affected by this disciplinary mismatch, although the matched anxiety findings remain subject to the other limitations of the secondary panel.

2.4. Quantitative Measures

All focal D1, D4, and D5 structured scale items used the same six observed response categories: 1 = completely inconsistent/not at all characteristic, 2 = somewhat inconsistent, 3 = relatively inconsistent, 4 = relatively consistent, 5 = very consistent, and 6 = completely consistent. No reverse-keyed item was identified in the implemented item map. Dimension scores were calculated as item means; higher scores indicated stronger endorsement of the corresponding literacy or anxiety content.

2.4.1. Generative AI Literacy

Generative AI literacy was assessed using a 16-item music-context-adapted self-report instrument derived from the four-domain framework of the Generative AI Literacy in Digital Multimodal Composing Scale (GAIDMCS) developed by Liu et al. [11]. The administered instrument contained four items each for Cognitive Generative AI Literacy (CL), Behavioral Generative AI Literacy (BL), Affective Generative AI Literacy (AL), and Ethical Generative AI Literacy (EL). It should not be interpreted as an exact shortened reproduction of the validated 17-item GAIDMCS.

The published GAIDMCS contains 17 items distributed across four dimensions (4 cognitive, 5 behavioral, 5 affective, and 3 ethical). A retrospective source-to-administered-item audit conducted for the revision showed that the present 4 × 4 instrument involved deletion of more than one source item: five source items had no direct counterpart in the focal 16-item instrument, four administered items had no sufficiently close counterpart among the 17 final GAIDMCS items, and several other correspondences involved either professional-context substitution or more substantial modification. Supplementary Table

S1A provides the item-level literacy audit; Supplementary Table S1C provides the complete map of the 59-item questionnaire battery.

In the implemented instrument, CL assessed self-reported understanding, application, analysis, and evaluation of GenAI in music-production and music-education contexts. BL assessed participation in GenAI-supported practice, routine use, willingness to invest effort in learning, and anticipated continued use. AL assessed perceived personalization, creativity enhancement, positive expectations, and perceived ease of use. EL assessed awareness of bias/discrimination, misuse and copyright-related risks, misinformation, and ethical implications for human creativity and professional expertise. No overall GenAI literacy total score served as a primary predictor.

Because the administered 16-item instrument differed from the source GAIDMCS in item allocation, wording, and disciplinary context, psychometric evidence from Liu et al. [11] was not assumed to transfer directly. Its measurement properties were evaluated locally through factor-model fit, composite reliability, average variance extracted, factor correlations, the heterotrait–monotrait ratio (HTMT), and an alternative model that merged CL and BL. No contemporaneous item-by-item adaptation log or independently documented translation/back-translation record was available; the source-to-adapted mapping in Supplementary Table S1A is therefore a retrospective documentary audit rather than a reconstruction of undocumented pre-study decisions.

2.4.2. Multidimensional AI Anxiety

AI Anxiety was assessed with the 21-item item family developed by Wang and Wang [17]: AI Learning Anxiety (LA; eight items), AI Job-Replacement Anxiety (JR; six items), Sociotechnical Blindness Anxiety (ST; four items), and AI Configuration Anxiety (AC; three items). LA concerned learning, understanding, operating, and keeping pace with AI. JR covered dependency, reduced reasoning, and human or job replacement. ST concerned misuse, malfunction, loss of control, and autonomous systems. AC concerned fear or discomfort regarding human-like AI-enabled technologies.

The original Artificial Intelligence Anxiety Scale used seven response categories, whereas the administered D1, D4, and D5 questionnaire used the six-category anchors reported above. A contemporaneous translation/adaptation record was not available. Local measurement evidence was therefore evaluated rather than inherited from the source instrument, and no overall AI Anxiety total score served as a primary outcome. A retrospective item-by-item correspondence audit showed close semantic correspondence between the published 21 AIAS items and the administered Chinese items, but this does not substitute for a documented translation procedure. Supplementary Table S1B reports the source-item content, administered wording, and response-format difference item by item.

2.5. Qualitative Sources

Three qualitative sources were analyzed. D9 contained 56 classroom-practice reflections from 55 unique identifiers in a Broadcasting/Hosting context. Four text prompts addressed AI functions, connections with previous experience, uncertainties, and the perceived role of AI in class. D11 contained 28 reflections from a first AI-video assignment, with six text prompts; its final prompt asked about video-production anxiety rather than the validated LA construct. D9 and D11 identifiers linked to other hosting questionnaire subsets but not to D1 or the 47 exact D4–D5 matches.

Q1 comprised three semi-structured interview sessions, four apparent student speakers, and 100 extracted student turns. The interviews addressed experience, learning, confidence, tools, ethics, creativity, traditional teaching, future use, anxiety, and difficulty. Speaker labels reset across sessions, so the material was cohort-linked rather than participant-linked.

Exact interview dates and D4–D5 linkage were unavailable. Uploaded dialogue-image URLs embedded in D9 were not opened or coded because of privacy, expiry, and third-party-content risks.

Before submitting D9 and D11 reflections, students were informed through the questionnaire/informed-consent page that their reflective material could be used for research. Permission covered publication of anonymized excerpts from D9, D11, and Q1. All retained quotations use de-identified source labels. The English excerpts used in the manuscript were checked against the original Chinese material and approved in their final wording; Q1 transcripts were also checked against corresponding audio recordings by Xin Yong.

2.6. Procedural Timeline and D4–D5 Verification

D4 and D5 were treated as two ordered repeated-measurement waves surrounding researcher-documented AI-supported creative learning activities. The source files yielded 47 exact one-to-one participant-identifier matches, with no duplicate or ambiguous matched pairs. Gender and year codes agreed across the two waves, and the 59 common structured items retained the same relative ordering and substantively equivalent wording after removal of wave-specific question numbering. Four D5-only evaluation items were excluded from repeated-measures analyses.

The calendar chronology could not be verified with the same confidence as participant identity and item equivalence. Researcher records described D4 as a pre-activity assessment on 1 March 2026 and D5 as a post-activity assessment on 30 June 2026, with four 30-min instructional activities between them. However, the questionnaire-platform exports contained different submission timestamps: D4 records ranged from 19 May to 12 June 2026, whereas D5 records ranged from 16 June to 26 June 2026. For every matched participant, the D4 platform timestamp preceded the corresponding D5 timestamp.

Because contemporaneous session-level records sufficient to reconcile the researcher-reported dates with the platform timestamps were not available, March 1 and June 30 are not treated as verified questionnaire-administration dates and no exact exposure interval is inferred. The platform timestamps are reported as directly observed file metadata, whereas the broader designation of D4 as the earlier/pre-designated wave and D5 as the later/post-designated wave is based on questionnaire design and researcher documentation. Exact dates of the four instructional sessions are not reported. Supplementary Table S6 summarizes the chronology and context audit.

The D4–D5 data consequently support ordered matched repeated-measures description, not causal intervention-effect estimation. The unresolved calendar discrepancy, absence of independently verified session dates, context mismatch in the literacy items, small matched sample, and absence of a randomized comparison group all constrain interpretation.

2.7. Data-Quality Screening

The D1 audit examined missing identifiers and scale cells, impossible values, duplicate identifiers, completion time, response strings, within-person variability, modal-response concentration, and robust multivariate distance. No attention-check variable was available. No record had a missing identifier, a missing focal item response, or an impossible response value. Fourteen records were completed in fewer than 59 s, which was treated as a warning rather than an exclusion criterion.

The original questionnaire contained 59 structured six-point items: 16 focal GenAI literacy items, 22 items from four additional nonfocal blocks (anthropomorphic interaction, AI comfort, technical skills, and helpfulness), and 21 focal AI Anxiety items. The substantive analyses used the 37 focal literacy/anxiety items, whereas all 59 items were used for the whole-battery straightlining screen (Supplementary Table S1C).

Exact uniform responding across all 59 structured items occurred in 105 of the 460 eligible records (22.8%). The exact-straightline criterion had been specified before outcome interpretation as a response-quality sensitivity rule. To prioritize response quality in the revised inferential hierarchy, the resulting 355 records from 351 identifier clusters are used as the primary D1 inferential sample, while the full eligible 460 records from 455 clusters are retained as a sensitivity analysis. Low within-person variability, high modal concentration, rapid completion, and multivariate distance were not used as independent exclusions.

A deliberately sensitive broad warning algorithm flagged 393 records for at least one potential concern and 67 for no major concern. These categories do not classify 393 records as invalid. The inference therefore distinguishes the exact straightline rule from the broader diagnostic flags rather than treating all warnings as exclusion criteria.

D4–D5 analyses also examined exact 59-item straightlining. Seven matched participants were flagged at either wave; excluding those pairs yielded a secondary sensitivity sample of 40. For D11, seven of 28 long responses were completed unusually quickly. No near-duplicate text was detected, AI-assisted writing could not be confirmed, and no D11 response was removed; instead, the qualitative evidence-strength rating was reduced.

2.8. Quantitative Analysis

Analyses were conducted in Python 3.12 using SciPy 1.18.1, NumPy 2.5.3, scikit-learn 1.9.1, matplotlib 3.10.3 and semopy 2.2.2. Confirmatory factor analyses treated the six-category questionnaire items as approximately continuous and used normal-theory maximum-likelihood-with-Wishart estimation. Because an ordinal weighted least squares mean- and variance-adjusted (WLSMV) estimator was not implemented, the confirmatory factor analysis (CFA) results are interpreted as local, estimator-specific measurement evidence rather than definitive validation.

Measurement-model inference was anchored in the quality-screened D1 sample ($n = 355$). For GenAI literacy, the primary specification was a four-correlated-factor model representing CL, BL, AL, and EL. Given the observed CL–BL overlap, an alternative three-factor model in which the eight CL and BL items loaded on a single combined factor was estimated as a targeted discriminant-validity sensitivity analysis. The 21-item AI Anxiety instrument was evaluated using its four-correlated-factor specification. Model evaluation used chi-square, the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA) with 95% confidence intervals (CIs), and the standardized root mean square residual (SRMR), complemented by composite reliability (CR), average variance extracted (AVE), standardized loadings, factor correlations, and the heterotrait–monotrait ratio (HTMT). No residual correlations or cross-loadings were added after inspection of the results. The same measurement specifications were repeated in the full eligible sample ($n = 460$) as sensitivity analyses. Detailed measurement evidence is reported in Supplementary Table S4. In response to concern about possible semantic overlap between Ethical Literacy (EL) and Sociotechnical Blindness Anxiety (ST), a reviewer-requested exploratory construct-overlap diagnostic was also conducted in the primary $n = 355$ sample. This included Pearson item-level correlations, observed scale-mean correlations, HTMT, and a maximum-likelihood two-factor exploratory loading diagnostic with varimax rotation for the eight EL/ST items. The same item-level correlation summary was calculated for EL–JR. These diagnostics were descriptive sensitivity analyses and did not redefine the prespecified primary model (Supplementary Table S7).

2.8.1. Primary D1 Dimension-Level Model

The primary association analysis used the quality-screened sample of 355 records from 351 participant-identifier clusters. A restricted system of four linear regression equations with correlated outcome residuals was estimated using iterative feasible generalized least squares. The prespecified equations were LA regressed on CL, BL, and AL; AC regressed on AL; JR regressed on EL; and ST regressed on EL. This produced the two-sided focal coefficients CL-LA, BL-LA, AL-LA, AL-AC, EL-JR, and EL-ST.

Holm correction was applied across the same six focal coefficients. Participant-identifier-cluster-robust sandwich covariance estimates, with a finite-sample correction, addressed repeated identifiers. Unstandardized coefficients came from the observed-score system; standardized coefficients came from the same system refitted after z-standardizing the included dimension scores. Predictor correlations, variance inflation factors, tolerances, condition indices, and a joint CL-BL Wald test were used to evaluate multicollinearity. The nonprespecified cells of the complete 4×4 literacy-anxiety map remained exploratory and were not used to alter the prespecified six-test family.

2.8.2. Sensitivity Analyses

The principal response-quality sensitivity re-estimated the same restricted four-equation system in the full eligible D1 sample of 460 records from 455 identifier clusters, with the same six-test Holm family. Additional checks compared simultaneous and separate CL-BL specifications, model-based and participant-clustered standard errors, and earliest versus latest duplicate-record retention.

Because prior GenAI exposure could plausibly be related to both self-reported literacy and AI Anxiety, an additional covariate-adjusted sensitivity analysis included the two prior-use variables collected in the original D1 questionnaire: prior GenAI use for video production and prior GenAI use for music production. Each was entered as a categorical factor with no prior use as the reference and separate indicators for occasional use (1–2/month), frequent use (1–4/week), and very frequent use (≥ 5 /week). Both prior-use factors were entered into every equation of the same restricted system. The six prespecified literacy-anxiety coefficients remained the focal test family and were again Holm adjusted. Gender and year were reported descriptively but were not added post hoc to the prespecified primary model. Supplementary Table S3 reports the prior-use-adjusted results.

2.8.3. Secondary D4–D5 Matched Repeated-Measures Analysis

Paired *t* tests compared the two ordered D4 and D5 waves for CL, BL, AL, EL, LA, JR, ST, and AC among the 47 exact identifier matches. Holm correction was applied across the eight paired tests. Each result included earlier- and later-wave means (M) and standard deviations (SD), the later-minus-earlier mean difference, a 95% confidence interval, Cohen's *d*, the *t* statistic, raw *p* value, and Holm-adjusted *p* value. Wilcoxon signed-rank tests provided sensitivity analyses where difference-score normality was materially questionable, and a second sensitivity excluded seven pairs with exact straightlining at either wave.

To assess selection and matching risk, the 47 matched participants were compared at D4 with the 53 D4 records that had no exact D5 identifier match. This was treated as a file-level matching-risk analysis rather than a verified dropout analysis. Baseline CL, BL, AL, EL, LA, JR, ST, and AC were compared using Welch tests, pooled-standard-deviation standardized mean differences, and Holm correction across the eight tests. Gender, year, prior video-related GenAI use, and prior music-related GenAI use were compared descriptively and with exact contingency-table tests; Cramér's *V* summarized categorical effect size. Supplementary Table S5 reports these analyses.

Because the literacy instrument was not aligned with the documented professional context of the learning activities, no post hoc item-selection analysis was used to construct a more context-congruent literacy measure. The original four literacy dimension scores were retained for transparent descriptive comparison, but their interpretation was restricted to the administered music-framed self-report measures.

2.9. Qualitative Analysis

The qualitative analysis followed a hybrid deductive–inductive thematic approach. Broad sensitizing domains concerned competence, task support, uncertainty, creativity, ethics, professional concern, agency, and anxiety; they were starting points rather than fixed themes. Xin Yong, a native/proficient Chinese speaker, conducted the primary coding and analysis directly from the Chinese-language source materials. Microsoft Excel was used only to organize and index the corpus. Repeated reading, comparison across student responses, first-cycle coding with explicit definitions and inclusion/exclusion criteria, second-cycle thematic development, active examination of negative and deviant cases, and retained analytical memos supported the analysis.

Junjie Gavin Wu and Wei Wei subsequently reviewed the coding framework/codebook, final themes, and negative/deviant cases. Where interpretive differences arose, the final determination was reached through supervisory/second-researcher review. No second researcher independently recoded the complete corpus, and no coding-reliability coefficient was calculated. Quantitative and qualitative findings were initially developed separately before mixed-methods integration.

Xin Yong was both the course instructor and doctoral researcher for D9, D11, and Q1. This dual role and the possible influence of the teacher–student relationship on student accounts and interpretation were considered in a retained reflexive journal/memo.

Six themes were finalized before cross-strand integration: Contextualized Operational Competence, Efficiency-Enabled Reallocation of Attention, Calibrated Use of AI Outputs, Human-Led Creative Collaboration, Ethical and Professional Vigilance, and Task-Specific Anxiety Responses. Frequencies, when used, referred to unique responses rather than keyword counts. A dedicated negative-case analysis retained accounts of no learning, no concern, no AI use, greater confidence in AI than self, increased anxiety, transferred anxiety, no change, and ambiguity.

2.10. Mixed-Methods Integration Procedures

Integration was conducted only after the quantitative and qualitative analyses had been completed independently. A prespecified joint display paired quantitative findings with thematically relevant qualitative evidence while explicitly recording linkage level and evidentiary strength.

Cross-strand integration was conducted at the construct, theme, cohort, and study-context levels rather than the participant level. D9, D11, and Q1 were not verified as participant-linked to either the D1 sample or the D4–D5 matched panel. Accordingly, no quotation, qualitative case, or thematic pattern was used to explain an individual participant's questionnaire score, longitudinal change, or position on a quantitative association.

CONVERGENCE denotes consistency between independently analyzed strands at the same bounded construct or study-context level; unless explicit identifiers link the strands, it does not imply participant-level co-occurrence. COMPLEMENTARITY denotes contextual or thematic extension without independent confirmation of the quantitative relation. DIVERGENCE denotes different directions, scopes, or constructs across strands and does not imply contradiction within the same individuals. SILENCE denotes insufficient qualitative evidence to assess the relevant quantitative finding.

The joint display retained focal quantitative findings regardless of whether qualitative material converged with, complemented, diverged from, or remained silent about them. Qualitative evidence was not used to rescue unstable quantitative findings or to infer unmeasured mechanisms. Because the D4–D5 literacy items were music-framed whereas the qualitative material primarily concerned broadcasting, script generation, and video production, cross-strand comparisons involving the repeated-measures literacy results were classified as contextual complementarity rather than convergence. Integration category and evidentiary strength were treated as separate judgments. Strength reflected quantitative robustness, construct alignment, breadth across qualitative sources, treatment of negative/deviant cases, and participant-linkage level; no numerical integration score was calculated. Because no qualitative source was participant-linked to D1 or D4–D5, no integrated finding was classified as strong participant-level evidence. The complete rubric is provided in Supplementary Table S8.

2.11. Ethics, Consent, and Privacy Procedures

The study, formally approved under the title Generative AI Literacy and Multidimensional AI Anxiety among Chinese Arts University Students: Mixed-Methods Evidence of Differentiated Adaptation, was approved by the Ethics Review Committee, School of Media, Sichuan Conservatory of Music, on 10 January 2026. The committee assigned no formal approval number. The approved scope covered online questionnaires, D9 and D11 reflective journals, semi-structured interviews and audio recording, and D4–D5 repeated questionnaires linked by participant identifiers. No amendment or continuing review was required. All participants were undergraduates aged 18 years or above; no minors participated.

Electronic informed consent was obtained before online questionnaire participation. The opening page presented study information and required participants to affirm voluntary agreement before proceeding. Before interviews and audio recording, the purpose, recording procedure, risks and benefits, confidentiality, and participant rights were explained; recording began only after separate verbal agreement, and participants could decline recording without penalty. Before submitting D9 or D11 reflections, students were informed through the questionnaire/informed-consent page that the material could be used for research. Permission covered publication of anonymized excerpts from D9, D11, and Q1.

The signed ethics certificate uses the term “written informed consent.” The researcher and institutional record confirmed that the implemented electronic questionnaire consent and separate verbal interview/audio-recording consent procedures were covered by the approved protocol. The study therefore reports the procedures actually implemented and does not characterize them as paper-based written consent. Participation was voluntary; participation, refusal, or withdrawal did not affect grades, assessment, academic standing, or access to regular instruction.

Limited participant identifiers were collected when required for research administration or repeated-measure matching. Identifiers used for matching, direct identifiers, IP addresses, and upload links were removed from analytical outputs where applicable. IP addresses may have been recorded automatically by the survey platform but were not used as substantive variables. Dialogue-image URLs containing possible third-party material were excluded from coding. Only the principal investigator had access to raw research data, which were stored in password-protected files on a password-protected local computer. The data will be retained for five years after study completion and then permanently deleted. Identifiable raw data will not be made public.

3. Results

3.1. Participant Characteristics and Data-Quality Audit

D1 contained 460 technically eligible questionnaire records and 455 participant identifiers. Ten records belonged to five duplicate-identifier groups. No record had a missing identifier, a missing focal scale response, or a response outside the 1–6 range. Exact uniform responding across all 59 structured items occurred in 105 records (22.8%), yielding the primary quality-screened sample of 355 records from 351 identifier clusters. The full eligible sample was retained as a sensitivity analysis.

Other response-process diagnostics were deliberately treated as warnings rather than exclusion criteria. Fourteen records were completed in fewer than 59 s; 128 had within-person standard deviation ≤ 0.30 ; 166 used the same response category for at least 80% of structured items; and 243 were flagged by a sensitive multivariate-distance rule. In total, 393 records had at least one broad warning and 67 had none. These counts do not imply that 393 records were invalid.

Participant characteristics for the primary and full D1 samples are shown in Table 1. In the primary sample, 296 records (83.4%) indicated female gender and 59 (16.6%) male gender. First-year, second-year, third-year, and fourth-year records comprised 47.9%, 36.3%, 3.9%, and 11.8%, respectively. The four most frequent raw major labels were Music Education, Musicology, Music Performance, and Dance Performance. The complete raw free-text distribution is reported in Supplementary Table S2. No D1 age variable existed in the original questionnaire. Prior-use distributions are also shown in Table 1; all reported participant-characteristic variables had complete data.

For the secondary repeated-measures source, D4 contained 100 unique identifiers and D5 contained 66. Forty-seven identifiers matched exactly, 53 D4 identifiers had no D5 match, and 19 D5 identifiers had no D4 match. The 47 exact matches represented 47.0% of the D4 file. This is a file-level matching proportion rather than verified dropout from a defined cohort.

3.2. Descriptive and Measurement Results

Measurement-model inference was anchored in the quality-screened D1 sample (Table 2). For the 16-item GenAI literacy instrument, the four-correlated-factor model showed mixed fit, $\chi^2(98) = 395.90$, CFI = 0.949, TLI = 0.937, RMSEA = 0.093, 95% CI [0.081, 0.104], and SRMR = 0.045. Incremental and residual fit were comparatively favorable, whereas RMSEA remained above conventional benchmarks for close approximate fit. The model therefore provides partial rather than unqualified support for the intended four-domain structure.

CL-BL discriminant separation improved after quality screening but remained incomplete. In the primary sample, the CL-BL latent correlation was 0.825 and HTMT was 0.843, compared with 0.878 and 0.891 in the full eligible sample. Collapsing CL and BL into a single factor produced materially poorer primary-sample fit, $\chi^2(101) = 651.05$, CFI = 0.905, TLI = 0.887, RMSEA = 0.124, 95% CI [0.113, 0.135], and SRMR = 0.059. CL and BL were therefore retained separately, but their unique coefficients are interpreted cautiously. Primary-sample literacy CR ranged from 0.909 to 0.946 and AVE from 0.715 to 0.814.

The four-correlated-factor AI Anxiety model also showed mixed primary-sample fit, $\chi^2(183) = 790.80$, CFI = 0.937, TLI = 0.928, RMSEA = 0.097, 95% CI [0.089, 0.105], and SRMR = 0.052. Primary CR ranged from 0.939 to 0.977 and AVE from 0.724 to 0.919. JR-ST was the strongest remaining overlap (latent $r = 0.772$; HTMT = 0.780), lower than in the full sample ($r = 0.850$; HTMT = 0.856). JR and ST were retained separately because their item content and prespecified analytical roles differed.

Table 2. Measurement-model evidence in the primary quality-screened and full eligible D1 samples.

Key Evidence	SRMR	RMSEA [95% CI]	TLI	CFI	Chi-Square (df)	Sample	Model
CL-BL $r = 0.825$; HTMT = 0.843; CR = 0.909–0.946; AVE = 0.715–0.814	0.045	0.093 [0.081, 0.104]	0.937	0.949	395.90 (98)	Primary $n = 355$	GenAI literacy 4-factor
Materially poorer fit than 4-factor model	0.059	0.124 [0.113, 0.135]	0.887	0.905	651.05 (101)	Primary $n = 355$	GenAI literacy CL + BL merged 3-factor
CL-BL $r = 0.878$; HTMT = 0.891	0.040	0.102 [0.092, 0.111]	0.939	0.950	563.25 (98)	Full $n = 460$	GenAI literacy 4-factor
Materially poorer fit than 4-factor model	0.050	0.130 [0.121, 0.140]	0.901	0.916	886.24 (101)	Full $n = 460$	GenAI literacy CL + BL merged 3-factor
JR-ST $r = 0.772$; HTMT = 0.780; CR = 0.939–0.977; AVE = 0.724–0.919	0.052	0.097 [0.089, 0.105]	0.928	0.937	790.80 (183)	Primary $n = 355$	AI Anxiety 4-factor
JR-ST $r = 0.850$; HTMT = 0.856	0.045	0.103 [0.096, 0.110]	0.936	0.944	1067.89 (183)	Full $n = 460$	AI Anxiety 4-factor

Note. Six-category items were treated as approximately continuous and estimated with normal-theory maximum likelihood/Wishart. Primary interpretation is anchored in $n = 355$. The four-factor literacy model fits materially better than merging CL and BL, but residual CL-BL overlap remains. Detailed CR, AVE, factor correlations, HTMT, and loading ranges are reported in Supplementary Table S4. Reviewer-requested EL-ST/EL-JR item-content diagnostics are reported in Supplementary Table S7.

The full eligible sensitivity analyses reproduced the overall pattern. The literacy four-factor model yielded chi-square(98) = 563.25, CFI = 0.950, TLI = 0.939, RMSEA = 0.102, 95% CI [0.092, 0.111], and SRMR = 0.040; the AI Anxiety model yielded chi-square(183) = 1067.89, CFI = 0.944, TLI = 0.936, RMSEA = 0.103, 95% CI [0.096, 0.110], and SRMR = 0.045. Detailed fit, CR/AVE, factor correlations, HTMT, and loading ranges are reported in Supplementary Table S4.

A reviewer-requested item-level audit further examined whether the positive EL-ST and EL-JR associations could reflect shared item content. In the primary $n = 355$ sample, the 16 EL-by-ST cross-item Pearson correlations ranged from 0.194 to 0.402 (mean = 0.306), compared with within-EL correlations of 0.678–0.863 and within-ST correlations of 0.699–0.877. The observed EL-ST scale-mean correlation was $r = 0.363$ and HTMT was 0.389. A two-factor maximum-likelihood loading diagnostic with varimax rotation showed intended absolute loadings of 0.760–0.923 for EL and 0.772–0.915 for ST, while absolute cross-loadings were 0.099–0.219. EL-JR overlap was weaker (scale-mean $r = 0.285$; HTMT = 0.308; cross-item correlations = 0.052–0.374). These results are consistent with empirical separation, but they do not eliminate shared risk-awareness semantics as one contributor to the positive associations (Supplementary Table S7).

3.3. Primary Quality-Screened Dimension-Level Associations

The six prespecified associations were estimated in the primary quality-screened sample of 355 records from 351 identifier clusters (Table 3). CL was not uniquely associated with LA after simultaneous inclusion of BL and AL, $B = -0.123$, standard error (SE) = 0.070, beta = -0.132 , 95% CI [$-0.260, 0.014$], raw $p = 0.0792$, Holm $p = 0.3169$. BL was also unsupported, $B = 0.054$, SE = 0.096, beta = 0.056, 95% CI [$-0.134, 0.242$], raw $p = 0.5766$, Holm $p = 1.000$.

Table 3. Internally prespecified primary associations in the primary quality-screened analysis and full-sample sensitivity.

Interpretation	Holm <i>p</i>	Raw <i>p</i>	95% CI for B	Beta	SE	B	Sample	Association
Unsupported	0.3169	0.0792	[−0.260, 0.014]	−0.132	0.070	−0.123	Primary <i>n</i> = 355	CL-LA
Unsupported	1.000	0.5766	[−0.134, 0.242]	0.056	0.096	0.054	Primary <i>n</i> = 355	BL-LA
Unsupported in primary	0.4308	0.1436	[−0.043, 0.299]	0.132	0.087	0.128	Primary <i>n</i> = 355	AL-LA
Directionally unstable vs. full	1.000	0.7583	[−0.131, 0.095]	−0.017	0.058	−0.018	Primary <i>n</i> = 355	AL-AC
Supported; robust	0.0026	0.0005	[0.089, 0.319]	0.217	0.059	0.204	Primary <i>n</i> = 355	EL-JR
Supported; robust	<0.0001	<0.0001	[0.166, 0.395]	0.303	0.058	0.281	Primary <i>n</i> = 355	EL-ST
Sensitivity: unsupported	0.7128	0.3564	[−0.195, 0.070]	−0.063	0.068	−0.062	Full <i>n</i> = 460	CL-LA
Sensitivity: unsupported	0.7128	0.3747	[−0.106, 0.281]	0.087	0.099	0.087	Full <i>n</i> = 460	BL-LA
Sensitivity-only support	0.0405	0.0135	[0.046, 0.400]	0.216	0.090	0.223	Full <i>n</i> = 460	AL-LA
Sensitivity-only; direction changes	0.0293	0.0073	[0.041, 0.266]	0.139	0.057	0.154	Full <i>n</i> = 460	AL-AC
Supported in sensitivity	<0.0001	<0.0001	[0.233, 0.433]	0.343	0.051	0.333	Full <i>n</i> = 460	EL-JR
Supported in sensitivity	<0.0001	<0.0001	[0.299, 0.493]	0.415	0.050	0.396	Full <i>n</i> = 460	EL-ST

Note. Primary inference uses 355 records from 351 participant-identifier clusters after excluding 105 exact whole-battery straightliners. The full eligible 460-record/455-cluster sample is a sensitivity analysis. Participant-identifier-cluster-robust standard errors were used. Holm correction was applied to the same six prespecified coefficients within each sample. Robust denotes consistency in direction and adjusted inferential status across the sample specifications and does not imply causal or temporal direction. The retained internal analysis plan specified the six-test family before Phase 5 outcome testing; it was not externally preregistered.

Neither affective-literacy association met the primary Holm-adjusted criterion. AL-LA was $B = 0.128$, $SE = 0.087$, $\beta = 0.132$, 95% CI [−0.043, 0.299], raw $p = 0.1436$, Holm $p = 0.4308$. AL-AC was $B = -0.018$, $SE = 0.058$, $\beta = -0.017$, 95% CI [−0.131, 0.095], raw $p = 0.7583$, Holm $p = 1.000$.

Two associations met the primary Holm-adjusted criterion. EL was positively associated with JR, $B = 0.204$, $SE = 0.059$, $\beta = 0.217$, 95% CI [0.089, 0.319], raw $p = 0.0005$, Holm $p = 0.0026$, and with ST, $B = 0.281$, $SE = 0.058$, $\beta = 0.303$, 95% CI [0.166, 0.395], raw $p < 0.0001$, Holm $p < 0.0001$. These are cross-sectional associations and do not establish temporal or causal direction.

3.4. Full-Sample and Covariate-Adjusted Sensitivity Analyses

The full eligible sample of 460 records from 455 identifier clusters was used as a response-quality sensitivity analysis (Table 3). CL-LA and BL-LA remained unsupported. EL-JR remained positive, $B = 0.333$, $SE = 0.051$, $\beta = 0.343$, 95% CI [0.233, 0.433], Holm $p < 0.0001$, and EL-ST remained positive, $B = 0.396$, $SE = 0.050$, $\beta = 0.415$, 95% CI [0.299, 0.493], Holm $p < 0.0001$. By contrast, AL-LA, $B = 0.223$, $\beta = 0.216$, Holm $p = 0.0405$, and AL-AC, $B = 0.154$, $\beta = 0.139$, Holm $p = 0.0293$, met the adjusted criterion only in the full sample. AL-LA is therefore sample-specification sensitive, and AL-AC is additionally directionally unstable because its primary estimate was slightly negative.

Prior-use adjustment provided an additional check on the two Ethical Literacy findings. After simultaneous adjustment for prior video- and music-related GenAI use in the primary sample, EL remained positively associated with JR, $B = 0.224$, $SE = 0.057$, $\beta = 0.239$, 95% CI [0.113, 0.336], Holm $p = 0.0004$, and with ST, $B = 0.289$, $SE = 0.056$, $\beta = 0.311$, 95% CI [0.179, 0.398], Holm $p < 0.0001$. The same two associations remained positive in the full eligible adjusted analysis. Thus, the EL-JR and EL-ST estimates were not materially

attenuated by adjustment for the two measured prior-use indicators, although residual confounding cannot be excluded (Supplementary Table S3).

Other model-specification checks did not materially alter the direction of the EL-JR and EL-ST results. Nonprespecified associations from the complete 4×4 literacy-anxiety map remain exploratory and are not used to modify the six-test prespecified conclusions. Figure 2 displays the primary and full-sample standardized coefficients and their Holm-adjusted status.

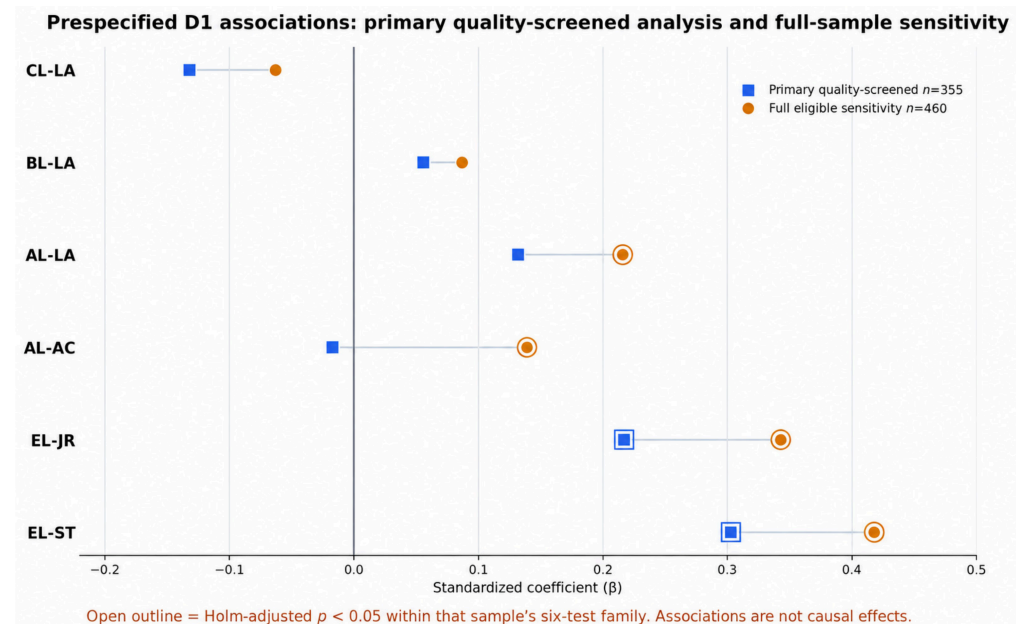


Figure 2. Internally prespecified primary D1 associations in the primary quality-screened analysis and full-sample sensitivity. Squares show standardized coefficients for the primary $n = 355$ sample and circles show the full eligible $n = 460$ sensitivity. Open markers indicate Holm-adjusted $p < 0.05$ within the same six-test family. EL-JR and EL-ST meet the adjusted criterion in both samples; AL-LA and AL-AC do so only in the full sample. Associations are not causal effects.

3.5. Secondary Matched Repeated-Measures Results

The D4–D5 analysis is secondary and requires two qualifications before the paired estimates are interpreted. First, the calendar-date discrepancy described in Section 2.6 remains unresolved, although every matched D4 platform timestamp preceded the corresponding D5 timestamp. Second, the literacy items were framed around music production/music learning whereas the documented learning activities centered on broadcasting scripts and AI-supported video production. The literacy comparisons therefore describe within-person differences in the administered music-framed self-report scores rather than validated activity-specific literacy gains.

Selection/matching-risk analyses compared the 47 exact matches with the 53 D4 records lacking an exact D5 match (Supplementary Table S5). None of the eight baseline dimension comparisons met the Holm-adjusted criterion. Standardized mean differences ranged from $|d| = 0.020$ to 0.288 ; the largest imbalance involved CL, for which the matched group had a somewhat higher D4 mean ($M = 3.463$, $SD = 0.872$) than the D4-only group ($M = 3.189$, $SD = 1.018$), $d = 0.288$, raw $p = 0.150$, Holm $p = 1.000$. Gender, year, prior video-related GenAI use, and prior music-related GenAI use also showed no statistically detectable distributional differences, although prior music-related use showed the largest categorical effect size (Cramér's $V = 0.178$). These results do not establish equivalence or absence of selection bias.

A response-quality difference also warranted caution: five of the 47 matched participants (10.6%) showed exact 59-item straightlining at D4, compared with none of the 53 D4-only records (Fisher's exact $p = 0.020$). Repeating the baseline dimension comparisons after excluding these five matched D4 straightliners did not produce any Holm-adjusted group difference.

Among the 47 exact matches, CL was higher at the later wave, $M = 4.176$, $SD = 0.773$, than at the earlier wave, $M = 3.463$, $SD = 0.872$. The later-minus-earlier difference was 0.713, 95% CI [0.449, 0.977], $dz = 0.792$, $t(46) = 5.43$, raw $p < 0.0001$, Holm $p < 0.0001$. BL was also higher later, $M = 4.027$ ($SD = 0.926$) versus 3.473 ($SD = 0.917$), difference = 0.553, 95% CI [0.307, 0.799], $dz = 0.661$, Holm $p = 0.0003$. AL was likewise higher later, $M = 4.197$ ($SD = 0.889$) versus 3.840 ($SD = 0.899$), difference = 0.356, 95% CI [0.151, 0.562], $dz = 0.510$, Holm $p = 0.0063$. These differences are statistically supported within the administered scales but are not interpreted as domain-aligned intervention gains.

EL was numerically higher at the later wave, $M = 4.394$ ($SD = 0.812$) versus 4.064 ($SD = 0.930$), but did not meet the eight-test Holm criterion, difference = 0.330, 95% CI [0.067, 0.592], $dz = 0.369$, raw $p = 0.0149$, Holm $p = 0.0747$. None of the four AI Anxiety dimensions showed a Holm-adjusted repeated-measures difference: LA difference = 0.226, Holm $p = 0.2833$; JR = -0.074 , Holm $p = 1.000$; ST = -0.122 , Holm $p = 0.8699$; and AC = -0.014 , Holm $p = 1.000$ (Table 4; Figure 3).

Table 4. D4–D5 matched repeated-measures comparisons.

Decision	Holm p	Raw p	$t(46)$	dz	95% CI	Later-Earlier Difference	Later M (SD)	Earlier M (SD)	Dimension
Higher at later wave	<0.0001	<0.0001	5.43	0.792	[0.449, 0.977]	0.713	4.176 (0.773)	3.463 (0.872)	CL
Higher at later wave	0.0003	<0.0001	4.53	0.661	[0.307, 0.799]	0.553	4.027 (0.926)	3.473 (0.917)	BL
Higher at later wave	0.0063	0.0011	3.50	0.510	[0.151, 0.562]	0.356	4.197 (0.889)	3.840 (0.899)	AL
No adjusted difference	0.0747	0.0149	2.53	0.369	[0.067, 0.592]	0.330	4.394 (0.812)	4.064 (0.930)	EL
No adjusted difference	0.2833	0.0708	1.85	0.270	[-0.020 , 0.472]	0.226	3.455 (1.083)	3.229 (0.879)	LA
No adjusted difference	1.000	0.5558	-0.59	-0.087	[-0.327 , 0.178]	-0.074	3.819 (0.988)	3.894 (0.809)	JR
No adjusted difference	0.8699	0.2900	-1.07	-0.156	[-0.352 , 0.108]	-0.122	4.000 (0.847)	4.122 (0.783)	ST
No adjusted difference	1.000	0.9110	-0.11	-0.016	[-0.268 , 0.240]	-0.014	3.326 (1.292)	3.340 (1.273)	AC

Note. $N = 47$ exact identifier matches. Holm correction covered eight paired t tests. For CL, BL, AL, and EL, the administered items were framed around music-production/music-learning contexts that did not match the documented broadcasting-script/video-production activities. Literacy differences should therefore be interpreted as changes in the administered music-framed self-report scores, not as validated activity-specific literacy development. The AI Anxiety items were domain-general. These are descriptive within-person differences and do not estimate intervention effects.

Wilcoxon signed-rank sensitivities preserved the central statistical pattern. Excluding seven pairs with exact straightlining at either wave left 40 pairs and again retained Holm-adjusted CL, BL, and AL differences only. These are observed within-person score differences across ordered waves, not estimates of causal instructional effectiveness.

D4–D5 matched ordered-wave dimension differences (N=47)

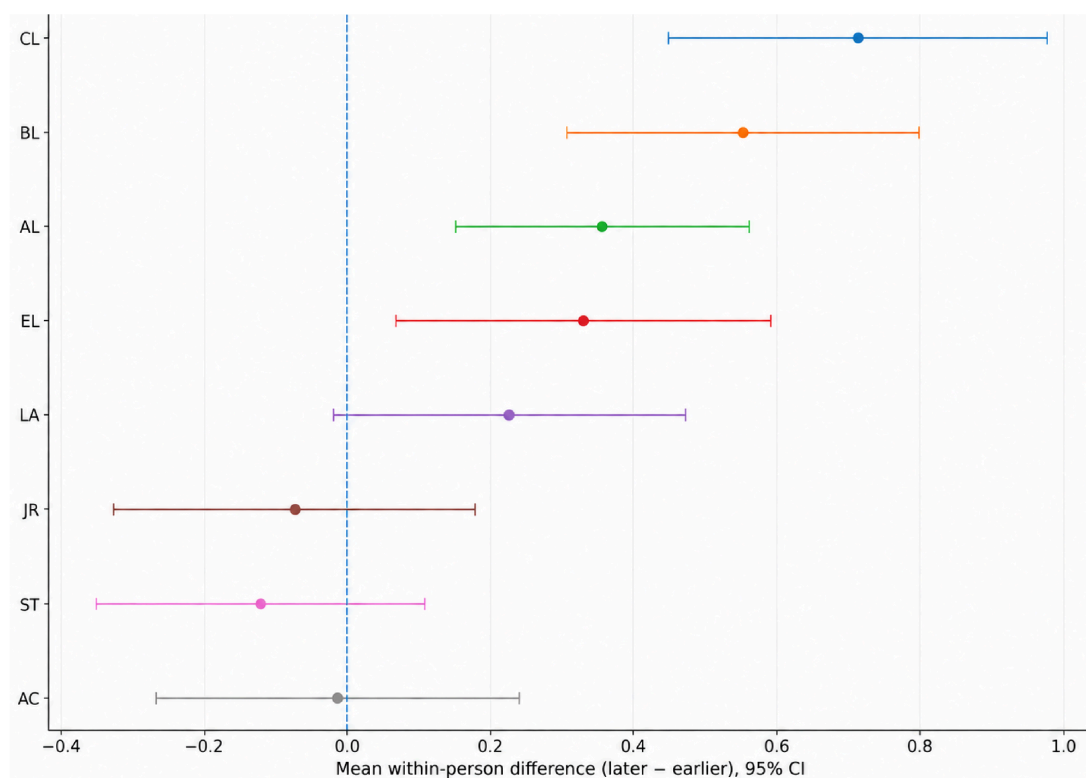


Figure 3. Matched D4–D5 ordered-wave within-person differences across literacy and anxiety dimensions. Points show later-minus-earlier mean differences and 95% confidence intervals for 47 exact identifier matches. CL, BL, and AL meet the eight-test Holm-adjusted criterion; EL, LA, JR, ST, and AC do not. The literacy dimensions were measured with music-framed items despite the broadcasting/video context of the documented activities; the plotted literacy differences are exploratory and should not be interpreted as activity-specific learning effects.

3.6. Qualitative Findings

3.6.1. Contextualized Operational Competence

Students described learning tool functions, improving prompt specificity, and connecting GenAI to discipline-related workflows. One D11 respondent wrote, “I learned techniques for using AI to generate video scripts” (D11-R008). A D9 respondent explained that they had previously not known a documentary-script structure but now understood how to complete one (D9-R036). Interview material similarly described becoming more adept through repeated adjustment. Competence was not universal or complete. One interview participant requested that the course “include more hands-on practice” (Q1-S1-P2-T062), while another D9 response simply stated, “I don’t know” (D9-R024).

3.6.2. Efficiency-Enabled Reallocation of Attention

AI was often described as reducing repetitive production labor and lowering entry barriers. Students linked this efficiency to attention that could be redirected toward content or expression. One D9 respondent wrote, “The assistant can help us reduce mechanical and repetitive work” (D9-R001). A D11 respondent stated that AI “allowed me to devote more energy to content creativity and broadcast expression” (D11-R001). These accounts concern perceived workflow reallocation rather than objectively measured productivity.

3.6.3. Calibrated Use of AI Outputs

Students did not uniformly treat generated content as ready for direct use. They described revising prompts, checking output, regenerating material, and retaining human oversight. One participant stated that AI output “cannot be used directly” (D9-R043). Another noted that the same visual could be attempted repeatedly without producing the intended result (D11-R022). An interview participant described AI as sometimes making unwanted decisions unless the prompt was constrained (Q1-S3-P3-T070). A minority reported no difficulty, which prevents treating output uncertainty as universal.

3.6.4. Human-Led Creative Collaboration

Students commonly positioned GenAI as an assistant while retaining authorship, aesthetic judgment, emotional expression, and professional decision-making as human responsibilities. One reflection stated, “The core of broadcasting is human warmth and thought; AI cannot replace it” (D9-R008). Another argued that “AI also cannot replace human aesthetics and emotional expression” (D11-R006). Some interview accounts nevertheless described AI as a source of inspiration. A deviant case judged the generated video potentially better than the student’s own on-camera work (D11-R025). The theme therefore reflects negotiated human leadership rather than unanimous human superiority.

3.6.5. Ethical and Professional Vigilance

Students raised questions about copyright, portrait rights, misinformation, algorithmic bias, authenticity, dependency, skill erosion, and future professional capability. One D11 respondent questioned “the authenticity of AI-generated content and whether algorithmic bias exists” (D11-R016). Another worried about being competent for future video-production work (D11-R026). Direct occupational-displacement language was comparatively sparse, and one interview participant reported having no concerns from beginning to end (Q1-S1-P2-T036). Ethical and professional vigilance was therefore present but not universal.

3.6.6. Task-Specific Anxiety Responses

D11 responses described heterogeneous feelings about concrete video-production tasks. Twenty of 28 usable reflections contained some description of reduced production or execution anxiety. One respondent wrote, “My anxiety decreased markedly” (D11-R001). This directional summary did not measure LA or any other validated AI Anxiety dimension.

Negative and deviant cases complicated a simple relief narrative. One student reported that technical-execution anxiety decreased while anxiety about creative control and outcome certainty increased substantially (D11-R017). Another reported a marked increase in video-production anxiety and questioned future competence (D11-R026). A third reported little meaningful feeling or change (D11-R028), while another had not used AI to make the video (D11-R003). Two responses were too ambiguous for directional interpretation. The task-specific pattern was therefore mixed rather than uniformly reassuring.

All English excerpts above were translated from Chinese, retain de-identified source identifiers, and use the final English wording manually verified and approved by the researcher.

3.7. Mixed-Methods Integration Findings

The joint display (Table 5) did not support a single overall change narrative, and all integration was interpreted at the construct or contextual level rather than as participant-level explanation.

Table 5. Revised mixed-methods joint display with explicit linkage boundaries.

Claim Strength	Integrated Interpretation	Integration	Linkage/Contextual Support	Qualitative Theme/Source IDs	Quantitative Finding/Status
Tentative–moderate	Operational-learning themes contextualize the later-wave CL difference but do not explain individual change.	COMPLEMENTARITY	Not participant-linked; context-misaligned	Contextualized Operational Competence Sources: D11-R008; D9-R036; Q1-S1-P2-T062	CL higher later wave—Secondary repeated-measures difference
Tentative–moderate	Iterative GenAI use contextualizes BL but does not validate domain-aligned development.	COMPLEMENTARITY	Not participant-linked; context-misaligned	Operational Competence; Calibrated Use Sources: D9-R043; D11-R022; Q1-S1-P2-T062	BL higher later wave—Secondary repeated-measures difference
Tentative	Some accounts describe greater ease, but affective responses are heterogeneous.	COMPLEMENTARITY	Tentative; not participant-linked	Operational Competence; Task-Specific Anxiety Sources: Q1-S3-P3-T065; D11-R001; D11-R026	AL higher later wave—Secondary repeated-measures difference
Presence moderate; change unsupported	Ethical issues are salient, but qualitative sources do not track EL change.	SILENCE	Moderate content; not linked to D4–D5	Ethical and Professional Vigilance Sources: D11-R016; Q1-S1-P2-T036	EL adjusted null—Secondary adjusted null
Tentative	Task-specific production emotion differs in scope from LA; no within-person contradiction is inferred.	DIVERGENCE	Different participants and construct	Task-Specific Anxiety Responses Sources: D11-R001; D11-R017; D11-R026	LA adjusted null—Secondary adjusted null
Low	Professional value is discussed, but longitudinal JR change is not tracked.	SILENCE	Low for temporal JR change	Human-Led Collaboration; Ethical/Professional Vigilance Sources: D11-R026; Q1-S1-P2-T036	JR adjusted null—Secondary adjusted null
Moderate context	Sociotechnical vigilance is salient without evidence of within-person ST persistence.	COMPLEMENTARITY	Contextual; not participant-linked	Ethical and Professional Vigilance; Calibrated Use Sources: D11-R016; D9-R043	ST adjusted null—Secondary adjusted null
Insufficient integrated evidence	Operational uncertainty does not explain or corroborate fear of human-like AI products.	SILENCE	Prompt/output-control examples concern a different construct	No construct-matched evidence in the cited examples Sources: None used to support AC	AC adjusted null—Secondary adjusted null

Table 5. Cont.

Claim Strength	Integrated Interpretation	Integration	Linkage/Contextual Support	Qualitative Theme/Source IDs	Quantitative Finding/Status
Strong quantitative; tentative integrated	Professional-value themes are compatible with the robust association but do not confirm it within individuals.	COMPLEMENTARITY	Not linked to D1; explicit displacement sparse	Ethical/Professional Vigilance; Human-Led Collaboration Sources: D11-R026; D9-R008; Q1-S1-P2-T036	EL-JR positive—Supported primary and sensitivity
Moderate contextual support; not participant-linked	Independent strands are consistent on ethical/sociotechnical concern; semantic adjacency may contribute, and no participant-level co-occurrence is established.	CONVERGENCE (construct level only)	Contextual; not linked to D1	Ethical/Professional Vigilance; Calibrated Use Sources: D11-R016; D9-R043; Q1-S3-P3-T070	EL-ST positive—Supported primary and sensitivity
Not core	Qualitative evidence does not validate or rescue the sensitivity-only coefficient.	DIVERGENCE	Different cohort/construct; heterogeneous	Task-Specific Anxiety Responses Sources: D11-R001; D11-R017; D11-R026; D11-R028	AL-LA full-sample positive—Not supported in primary
Not core	Qualitative material cannot resolve unstable quantitative direction.	SILENCE	Low for path direction; not linked to D1	No direction-equivalent qualitative evidence Sources: D11-R022; Q1-S3-P3-T070	AL-AC direction changes—Directionally unstable
Moderate and task-bounded	Task relief co-occurs at study level with broader anxiety nulls but involves different constructs/participants.	COMPLEMENTARITY	Prompt-bounded; not D4–D5	Task-Specific Anxiety Responses Sources: D11-R001; D11-R017; D11-R026; D11-R028	D11 task-specific relief—Qualitative only; 20/28 describe some reduction

Note. Qualitative sources were not verified as participant-linked to either D1 or the D4–D5 matched panel. All integration labels therefore refer to relationships between independently derived aggregate quantitative patterns and qualitative themes at the construct or study-context level. They do not indicate that the same individuals exhibited the quantitative and qualitative patterns. Representative quotations illustrate themes rather than explain individual questionnaire scores or trajectories. Integration labels and contextual-support strength are defined separately in Supplementary Table S8; no row represents participant-level mixed-methods convergence.

The higher later-wave CL and BL scores in the D4–D5 panel were accompanied by independently derived qualitative themes describing tool learning, prompt refinement, script/video production, output revision, and selective use of GenAI. However, the qualitative sources were not linked to the 47 matched participants, and the quantitative literacy items were music-framed whereas the qualitative materials primarily concerned broadcasting/video-production contexts. These relationships were therefore classified as contextual COMPLEMENTARITY rather than CONVERGENCE. The qualitative evidence illustrates broader forms of operational engagement with GenAI but does not explain the matched participants' score changes or validate the construct meaning of those changes.

The higher later-wave AL score received only tentative complementarity from accounts describing ease, willingness, confidence, frustration, and uncertainty. For EL, the absence of an adjusted D4–D5 difference was accompanied by qualitative evidence that ethical and professional issues were salient, but this was classified as SILENCE regarding longitudinal change because the qualitative sources did not track the matched panel. Qualitative discussion of sociotechnical risks provided contextual complementarity for ST without establishing temporal persistence or within-person correspondence. Prompting and output-control difficulties concern a different construct from AC, which measures fear of human-like AI products. The cited qualitative material therefore provides SILENCE regarding the AC finding.

The most direct construct-level cross-strand alignment concerned the D1 EL-ST association. EL-ST remained statistically supported across the response-quality specifications, while independently analyzed qualitative material repeatedly addressed misinformation, bias, authenticity, rights, control, reliability, and responsibility. This was classified as construct-level CONVERGENCE: the quantitative association and qualitative theme were compatible with a study-level interpretation in which ethical attention and sociotechnical concern were jointly salient. Because the qualitative sources were not linked to D1 participants, this does not demonstrate that the same individuals with higher EL scores were those articulating sociotechnical concern.

EL-JR showed weaker cross-strand alignment. The quantitative association was robust, but explicit occupational-displacement language was sparse in the qualitative material, so the relationship was classified as tentative COMPLEMENTARITY. The nonrobust AL-LA association and heterogeneous D11 task-specific anxiety reports were classified as DIVERGENCE in construct-level pattern and scope, not as contradictory responses from the same participants. The unstable AL-AC association remained qualitatively unresolved.

4. Discussion

4.1. Principal Findings

The primary D1 inference is now anchored in the quality-screened sample rather than the full eligible sample. Only two of the six prespecified associations met the Holm-adjusted criterion in the primary analysis: EL was positively associated with both JR and ST. Both relations remained positive in the full eligible sample and after adjustment for measured prior video- and music-related GenAI use. CL-LA and BL-LA were unsupported in both sample specifications, whereas AL-LA and AL-AC were sample-sensitive and are not treated as core findings.

The secondary D4–D5 panel showed higher later-wave CL, BL, and AL scores but no Holm-adjusted difference in EL or any AI Anxiety dimension. These literacy differences are exploratory because the administered items were music-framed while the documented activities involved broadcasting scripts and video generation; the paired design also lacks a comparison group and has unresolved chronology and selection limitations. Independently analyzed qualitative material described operational learning, selective revision, human-led

collaboration, ethical vigilance, and heterogeneous task-specific anxiety, but was integrated only at construct/theme/context level.

Together, the findings support differentiated rather than uniform adaptation as a bounded study-level organizing proposition. They do not show that higher literacy reduces anxiety, that ethical literacy produces concern, or that the documented learning activities caused the repeated-measures differences.

4.2. Differentiated Rather than Uniform Psychological Adaptation

Recent evidence increasingly suggests that AI-related capability and anxiety should not be interpreted through a single buffering model. In the present primary D1 analysis, EL-ST and EL-JR were supported, CL-LA and BL-LA were not, and the affective-literacy relations were sensitive to sample specification. The secondary repeated-measures results were also nonuniform: three music-framed literacy scores were higher at the later wave, while the four AI Anxiety dimensions did not meet the adjusted criterion. Because the strands involve different samples and different evidentiary strengths, this pattern is interpreted at the study level rather than as a single learner trajectory.

Broader educational-psychology research helps explain why a nonuniform pattern is plausible. Wang et al. [19] showed that relations among anxiety, attention, and performance varied with emotion regulation and state motivation; Putwain et al. [20] found reciprocal relations between test anxiety and achievement; and Taboada Barber et al. [21] located reading anxiety within a broader system of self-efficacy, cognitive processes, engagement, and achievement. These are not AI-anxiety studies, but they support caution against assigning one psychological meaning to anxiety scores.

AI-specific evidence points in the same direction without implying that anxiety is beneficial. Song et al. [25] found that GenAI assistance could improve creative performance while increasing fear-related emotion in one condition, and Jou and Hao [29] treated AI substitution as a psychologically salient learning event within a scaffolded career-reflection process. The present data therefore support only a bounded proposition: practical capability, affective orientation, ethical attention, and different forms of AI-related concern need not align in one direction across evidence strands.

4.3. Later-Wave Literacy Differences Without Uniform Anxiety Reduction

The D4–D5 panel provides secondary evidence of nonuniform score change. CL, BL, and AL were higher at the later wave, whereas EL and all four AI Anxiety dimensions did not meet the adjusted criterion. Independently analyzed qualitative material described tool learning, prompt refinement, iterative production, and selective output revision. Because these sources were not participant-linked to the matched panel and were not fully aligned with the music-framed literacy measures, they provide contextual complementarity rather than confirmation of literacy development.

The repeated-measures results therefore establish later-wave differences in the administered self-report scores, not participant-level evidence that the documented activities produced the qualitative experiences described in D9, D11, or Q1. The single-group design also cannot separate concurrent learning, repeated testing, maturation, regression to the mean, or selection from any activity-related contribution.

Intervention studies offer useful comparators without validating the present repeated-measures interpretation. Hsu et al. [24] found improved computational-thinking performance without significant between-group differences in AI concepts or AI anxiety. Connell Pensky et al. [16] reported gains in several GenAI competencies without equivalent improvement in critical bias evaluation. Jaggy et al. [15] likewise showed that exposure does not guarantee calibrated understanding. These studies make uneven change plausible,

but they do not convert the present music-framed literacy differences into activity-specific learning effects.

The absence of Holm-adjusted LA, JR, ST, or AC change should not be treated as psychological stasis. D11 task-specific anxiety reports were heterogeneous and concerned a concrete production task rather than the validated anxiety dimensions. Because D11 involved different participants and constructs, those accounts contextualize the study but cannot explain the D4–D5 scale results.

4.4. Ethical Literacy and Sociotechnical Vigilance

EL-ST was the most direct construct-level cross-strand alignment in the study, but its interpretation requires explicit consideration of construct-content overlap. Quantitatively, higher EL scores were associated with higher ST scores in the primary quality-screened sample and the full-sample sensitivity. EL and ST were empirically distinguishable in the primary sample, yet both contained risk-awareness content, particularly around misuse and adverse consequences. The reviewer-requested item-level diagnostics showed that EL-ST cross-item correlations were materially lower than within-construct correlations and that an exploratory two-factor diagnostic retained a clear intended-factor pattern. These results argue against treating EL and ST as interchangeable measures, but they do not exclude shared semantic content as one contributor to their association.

The most defensible interpretation is therefore narrower than a claim that EL-ST directly represents heightened critical vigilance: higher self-reported ethical attention was associated with greater sociotechnical concern across two conceptually adjacent but empirically separable domains. Independently analyzed qualitative material also foregrounded misinformation, algorithmic bias, authenticity, rights, control, reliability, and responsibility, providing construct-level contextual consistency rather than participant-level confirmation. The qualitative material cannot identify which D1 respondents expressed those concerns or explain why the EL-ST coefficient arose.

The cross-sectional design cannot determine whether ethical engagement heightened sociotechnical concern, pre-existing concern motivated ethical engagement, shared risk-oriented wording contributed to covariance, or both constructs were shaped by prior exposure, attitudes, or response style. The result therefore supports an association robust across the reported sensitivity analyses and construct-level alignment, not a claim that ethical literacy causes vigilance or anxiety.

A further implication concerns calibrated reliance. Hou et al. [30] distinguished reflective, cautious, thoughtless, and collaborative forms of GenAI reliance. The present qualitative theme of Calibrated Use of AI Outputs is conceptually compatible with that distinction, but it remains an interpretive theme rather than a measured quantitative construct. Educationally, the goal is not simply to reduce ST but to support proportionate attention to reliability, bias, control, rights, and responsibility alongside competent use.

4.5. Job-Replacement Concern with Tentative Qualitative Complementarity

EL was also positively associated with JR in the primary and sensitivity analyses, but the cross-strand evidence was weaker than for EL-ST. In the independently analyzed qualitative material, participants discussed future capability, human professional value, skill erosion, and the limits of AI, yet explicit occupational-displacement language was comparatively sparse. EL-JR is therefore treated as a robust quantitative association with tentative qualitative complementarity rather than as strong mixed-methods convergence. The reviewer-requested overlap audit also indicated weaker EL-JR content covariance than EL-ST (observed scale-mean $r = 0.285$; HTMT = 0.308), which reduces but does not eliminate concern about shared evaluative wording.

This distinction matters because the JR scale encompasses replacement, dependency, and reduced reasoning, whereas many qualitative comments concerned broader professional preparedness or the continuing value of human judgment. The evidence does not justify translating JR into an unmeasured construct such as professional identity threat. Jou and Hao [29] provide a relevant comparison concerning occupationally salient AI concerns in an active learning context. Hsu et al. [24] found no significant between-group difference in AI anxiety. Neither study directly validates the present EL-JR relation.

Several directional explanations remain plausible. Students already concerned about occupational change may attend more closely to ethical/professional consequences; greater critical engagement may make replacement and dependency more salient; or both may reflect shared exposure or response tendencies. The qualitative material contextualizes these possibilities but cannot be assigned to the D1 respondents contributing to the EL-JR coefficient.

4.6. Calibrated AI Use and Human-Led Creative Collaboration

The qualitative accounts portrayed GenAI as useful but provisional. Participants described generating scripts, images, or videos, revising prompts, checking factual or stylistic fit, regenerating unsatisfactory material, and retaining human oversight. Efficiency was valued when it released attention for content and expression, but participants also reported factual uncertainty, repeated generation, and unwanted system decisions. Hou et al. [30] offer a useful comparison because their reliance framework distinguished reflective, cautious, thoughtless, and collaborative GenAI use.

These themes should remain interpretive rather than be promoted into unmeasured quantitative constructs. Their value lies in illustrating forms of GenAI practice that are conceptually compatible with the quantitative patterns. The qualitative material shows examples of output checking, selective revision, human-led collaboration, and ethical/professional vigilance, but it does not establish that the same participants who scored higher on literacy were those who displayed these practices or concerns.

Accordingly, calibrated AI use remains a pedagogical proposition derived from cross-strand pattern alignment, not a demonstrated within-person developmental trajectory. The strongest instructional implication is to combine hands-on production with explicit output checking, comparison of human and AI contributions, discussion of authorship and rights, and reflection on when reliance is justified, rather than to maximize use or minimize concern.

4.7. Task-Specific Anxiety Versus Multidimensional AI Anxiety

Twenty of 28 D11 reflections described some reduction in production or execution anxiety. This qualitative pattern differed from the aggregate D4–D5 LA result, but the two should not be treated as directly discrepant evidence because they involved different participants and different constructs. LA measured broader anxiety about learning and operating AI, whereas D11 asked about a concrete video-production experience.

Negative and deviant cases further bounded interpretation. Some participants reported reduced technical burden but greater anxiety about creative control or outcome certainty; others reported increased anxiety, no change, ambiguity, or no AI use. The qualitative evidence therefore illustrates task-specific heterogeneity rather than generalized reassurance.

Seven D11 long responses were completed unusually quickly. Near-duplicate text was not detected and AI-assisted writing could not be confirmed, so no response was excluded; instead, evidentiary strength was downgraded. These records cannot establish prevalence beyond the analyzed prompt or substitute for validated LA measurement.

4.8. Implications for Educational Psychology

The educational-psychology contribution can be stated in three bounded propositions. First, dimension specificity: GenAI literacy and AI Anxiety should be modeled as differentiated systems rather than collapsed into total scores. The primary D1 results make this point empirically because EL showed supported relations with ST and JR, CL-LA and BL-LA were unsupported, and the AL relations were sample-sensitive. An overall literacy-anxiety coefficient could conceal this heterogeneity.

Second, competence-concern decoupling is used here only as a study-level non-lockstep pattern, not as statistical independence or a demonstrated within-person mechanism. The secondary panel showed higher later-wave scores on three administered literacy dimensions without Holm-adjusted change in the broader AI Anxiety dimensions. Because the literacy items were context-misaligned with the documented activities and the qualitative strands were not participant-linked, the result should not be read as proof that the same individuals gained domain-specific competence while retaining concern. It instead motivates separate measurement of capability and concern in stronger longitudinal designs.

Third, calibrated rather than anxiety-free adaptation is an educational proposition rather than a validated psychological construct. The EL-ST association and qualitative themes of output checking, human oversight, and ethical/professional vigilance indicate that critical attention can remain salient alongside GenAI use at the study level. The instructional implication is to combine competence building with verification, ethical reasoning, uncertainty management, and reflection on human agency rather than treating anxiety reduction as the sole endpoint of successful AI learning.

4.9. Implications for Arts and Creative Education

Creative tasks make output ownership, authenticity, emotional expression, and professional value visible. The documented script and video activities offered opportunities to practice prompting and production while discussing copyright, bias, creativity, and agency. Similar combinations have appeared in digital storytelling, vocal music, collaborative music creation, abstract art, and design education [2–5,31]. Recent evidence also points to heterogeneous AI-related responses across higher education and creative-arts contexts, including associations among AI literacy, attitudes, and self-efficacy in university students [32], AI-use intentions among art and design students [33], and replacement-related employment anxiety among vocal-music students [34].

The arts context should not be treated as a novelty label, and the present data do not support latent mean comparisons across disciplines. Its theoretical value lies in making authorship, expression, and human judgment concrete. Future work should use context-aligned measures and establish scalar or defensible partial scalar invariance before comparing creative disciplines or trajectories.

4.10. Methodological Implications

The revised analyses illustrate why response quality, dimensional structure, and inferential hierarchy matter. Exact whole-battery straightlining materially affected the AL findings and inflated apparent internal consistency and construct overlap. Anchoring inference in the quality-screened sample changed the interpretation from an overall unstable model to a more specific pattern: EL-JR and EL-ST were supported, CL-LA and BL-LA were unsupported, and the AL paths were sensitivity-dependent.

The measurement audit also shows why adaptation of an established scale should not be described as a simple shortened version when item allocation and content have changed. The 16-item literacy instrument requires local validation, and the mixed CFA fit and residual

CL-BL overlap bound interpretation even though the four-factor model fit materially better than a CL-BL merged alternative.

Finally, mixed-methods integration depends on preserving linkage and construct boundaries. Task-specific production anxiety, broadcasting/video workflow themes, and music-framed literacy scores can inform one another only at the construct/theme/context level in the present data. Silence, divergence, and complementarity are informative outcomes because they prevent qualitative material from becoming post hoc confirmation of unstable or unlinked quantitative findings.

4.11. Limitations

Several limitations materially bound the conclusions. First, D1 response quality was a major concern: 105 of 460 eligible records showed exact whole-battery straightlining. The quality-screened $n = 355$ sample is therefore used for primary inference, with $n = 460$ retained as sensitivity. EL-JR and EL-ST were supported in both, but the AL findings were sample-sensitive. The broad warning algorithm was intentionally sensitive and should not be interpreted as classifying all flagged records as invalid. For EL-ST specifically, item-level diagnostics supported empirical separation but also showed nontrivial cross-item correlations; shared risk-awareness semantics therefore remains a plausible contributor and the association is not interpreted as a demonstrated vigilance mechanism.

Second, all focal measures were self-reported and the 16-item GenAI literacy instrument was music-context-adapted rather than an exact reproduction of the validated 17-item GAIDMCS of Liu et al. [11]. A retrospective item audit showed omitted, newly added, contextualized, and substantially modified content. No contemporaneous item-by-item adaptation log or independently documented translation/back-translation record was available. Both primary CFA models showed mixed fit, and CL-BL overlap remained substantial despite improvement after quality screening. Six-category items were treated as approximately continuous rather than estimated with an ordinal WLSMV model.

Third, the D1 associations are cross-sectional and cannot establish direction or mechanism. Prior video- and music-related GenAI use could be adjusted because those variables were recoverable from the original questionnaire, but residual and unmeasured confounding remain possible. Major was a free-text field with ambiguous variants, and gender/year were not added post hoc to the prespecified primary model.

Fourth, the D4–D5 panel included only 47 exact matches and lacked a randomized or adequate nonrandomized comparison group. The matched panel represented 47% of the D4 file. Baseline comparisons did not identify Holm-adjusted group differences, but CL showed the largest standardized imbalance ($d = 0.288$), and five matched participants versus no D4-only records showed exact D4 straightlining. These diagnostics do not establish representativeness or absence of selection bias.

Fifth, procedural provenance and contextual validity limit the repeated-measures interpretation. Researcher-reported March 1/June 30 dates conflict with platform submission windows, and exact session dates could not be independently reconstructed. More importantly, the D4–D5 literacy items were framed around music production/music learning whereas the documented activities involved broadcasting scripts and AI-supported video production. The literacy differences therefore represent exploratory changes in the administered scores rather than validated domain-aligned literacy development. This context mismatch is less direct for the domain-general AI Anxiety items.

Sixth, participant-level linkage between the quantitative and qualitative strands was unavailable. D9, D11, and Q1 can contextualize constructs and study settings but cannot explain change within the 47 matched participants or identify which D1 respondents occupied particular positions on EL-ST or EL-JR. D11 production anxiety was not equivalent

to LA, and seven long D11 responses were completed unusually quickly. Finally, the primary analyst's dual instructor-researcher role and the single-institution setting may limit candor, interpretation, and transferability.

4.12. Future Research

Future research should prospectively document scale translation and contextual adaptation, validate context-appropriate literacy and anxiety measures with ordinal estimators, and pair self-report literacy with performance-based tasks. At least three measurement occasions would permit reciprocal or lagged analyses among CL, BL, AL, EL, LA, JR, ST, and AC rather than assuming one-way literacy-to-anxiety pathways. Cross-disciplinary comparisons should use harmonized wording and establish scalar or defensible partial scalar invariance before comparing latent means or trajectories.

Competence-concern decoupling should be tested as a developmental pattern rather than inferred from cross-sectional associations or a two-wave single-group panel. Controlled longitudinal studies should prospectively document attendance, exposure, timestamps, and concurrent AI use, and should align measurement wording with the actual instructional domain. Such designs could distinguish short-term task relief from durable changes in broader AI Anxiety and separate career awareness or perceived substitutability from JR.

Finally, testing calibrated rather than anxiety-free adaptation requires participant-linked mixed-methods evidence. Consented identifiers should connect surveys, reflections, interviews, performance tasks, and activity traces so that researchers can examine whether the same individuals who become more capable also become more selective in their reliance, more accurate in judging AI limitations, or more attentive to ethical and sociotechnical risk. Negative cases, including persistent concern, overreliance, disengagement, and increasing anxiety, should be retained rather than defining adaptation by declining anxiety alone.

5. Conclusions

This study found differentiated rather than uniform patterns across GenAI literacy and multidimensional AI Anxiety in the examined arts-university samples. In the primary quality-screened D1 analysis, EL remained positively associated with ST and JR, and both relations remained supported in the full eligible and measured-prior-use-adjusted sensitivity analyses. CL-LA and BL-LA were unsupported in both D1 samples, whereas AL-LA and AL-AC were sample-sensitive and are not treated as core findings.

In the secondary 47-person matched panel, CL, BL, and AL scores were higher at the later wave, whereas EL and all four AI Anxiety dimensions showed no Holm-adjusted difference. These literacy results are exploratory because the administered items were music-framed while the documented activities involved broadcasting scripts and video generation, and the design does not establish causal instructional effects.

Across evidence strands, independently derived qualitative themes concerning output checking, selective revision, human-led collaboration, and ethical/professional vigilance were compatible with the broader quantitative pattern, with the strongest construct-level convergence for EL-ST and only tentative complementarity for EL-JR. Because the qualitative sources were not participant-linked, the findings do not establish that the same individuals who showed higher literacy were also those expressing greater vigilance or selective reliance.

Accordingly, calibrated rather than anxiety-free adaptation is retained as a bounded study-level interpretive proposition, not as an observed within-person developmental trajectory. Stronger controlled longitudinal designs, context-aligned measurement, performance-based outcomes, and prospectively participant-linked mixed methods are needed to test that proposition.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ai7100417/s1>, Supplementary Tables S1A–C and S2–S8 provide the source-to-adapted item maps, complete 59-item block map, raw major distribution, prior-use-adjusted sensitivity analyses, measurement-model diagnostics, D4–D5 selection/matching-risk audit, chronology/context audit, EL-ST/EL-JR construct-content diagnostics, and the mixed-methods integration rubric referenced in the text.

Author Contributions: X.Y.: Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Visualization, Writing—Original Draft, Writing—Review & Editing, Project Administration. J.G.W.: Conceptualization, Methodology, Supervision, Writing—Review & Editing. W.W.: Supervision, Validation, Writing—Review & Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was approved by the Ethics Review Committee, School of Media, Sichuan Conservatory of Music, on 10 January 2026; no formal approval number was assigned. The approved study title was Generative AI Literacy and Multidimensional AI Anxiety among Chinese Arts University Students: Mixed-Methods Evidence of Differentiated Adaptation. Approval covered online questionnaires, reflective journals, semi-structured interviews and audio recording, and linked repeated questionnaires. All participants were undergraduates aged 18 years or above. The implemented consent procedures were confirmed as covered by the approved protocol.

Informed Consent Statement: Electronic informed consent was obtained before online questionnaires, and separate verbal informed consent was obtained before interviews and audio recording. These procedures were covered by the approved protocol despite the certificate's general reference to written informed consent. Students were informed that submitted reflections could be used for research. Permission covered publication of anonymized excerpts. Participation, refusal, or withdrawal did not affect grades, assessment, academic standing, or access to regular instruction.

Data Availability Statement: Selected de-identified questionnaire data, adequately de-identified eligible reflective-journal materials, reproducible analysis code, table and figure source data, and eligible supplementary or audit materials may be made available to qualified researchers upon reasonable request to Liu Li (sccmcmxy@163.com). Release is subject to review and final approval by Junjie Gavin Wu (gavinjunjiewu@gmail.com), de-identification and re-identification-risk checks, and applicable privacy and confidentiality restrictions. Full interview transcripts, interview audio recordings, identifiable raw participant data, and material not authorized for redistribution will not be shared. No public repository deposition is planned.

Acknowledgments: The authors declare that Generative AI was used in the creation of this manuscript. During the research and manuscript preparation process, ChatGPT 5.6 and OpenAI tools were used to support selected literature-review, methodological, computational, qualitative-analysis, mixed-methods integration, interpretation, scientific writing, visualization, and language-refinement tasks. The authors reviewed, edited, and verified all AI-assisted output incorporated into the study and take full responsibility for the accuracy, originality, interpretation, citations, and final content of the manuscript. The study design and data collection were conducted by the authors, while all final analytical decisions, qualitative coding and themes, interpretation of results, and final scholarly claims were determined and approved by the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xu, Y.; He, K.; Levine, J.; Ritchie, D.; Pan, Z.; Bustamante, A.; Warschauer, M. Artificial intelligence enhances children's science learning from television shows. *J. Educ. Psychol.* **2024**, *116*, 1071–1092. [[CrossRef](#)]
2. Hıçyılmaz, Y. An innovative approach in arts education: Student experiences of abstract art practices supported by generative artificial intelligence. *SAGE Open* **2025**, *15*, 21582440251382812. [[CrossRef](#)]
3. Li, W.; Cui, X.; Manoharan, P.; Dai, L.; Liu, K.; Huang, L. AI-assisted feedback and reflection in vocal music training: Effects on metacognition and singing performance. *Front. Psychol.* **2025**, *16*, 1598867. [[CrossRef](#)] [[PubMed](#)]

4. Song, F.; Xu, B. Effects of integrating a structured design thinking strategy into generative AI-supported design learning on students' design achievement, creative self-efficacy, and problem-solving skills. *Front. Psychol.* **2026**, *17*, 1847432. [[CrossRef](#)] [[PubMed](#)]
5. Wei, X.; Wang, L.; Lee, L.-K.; Liu, R. The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study. *Int. J. Educ. Technol. High. Educ.* **2025**, *22*, 23. [[CrossRef](#)]
6. Yan, W.; Nakajima, T.; Sawada, R. Beyond tool use: Tracking the evolution of generative AI literacy among university students through a process-oriented investigation. *Comput. Educ. Artif. Intell.* **2025**, *9*, 100465. [[CrossRef](#)]
7. Cao, S.; Liu, X.; Yu, X. AI literacy and high school English teachers' teaching anxiety: The chain mediation role of teaching self-efficacy and work engagement. *Front. Psychol.* **2026**, *17*, 1797047. [[CrossRef](#)] [[PubMed](#)]
8. Chen, C.; Hu, W.; Wei, X. From anxiety to action: Exploring the impact of artificial intelligence anxiety and artificial intelligence self-efficacy on motivated learning of undergraduate students. *Interact. Learn. Environ.* **2025**, *33*, 3162–3177. [[CrossRef](#)]
9. Teke, B. Can artificial intelligence anxiety serve as a facilitator: The relationship between pre-service teachers' AI anxiety, their motivated learning and AI self-efficacy. *Interact. Learn. Environ.* **2026**, *34*, 3086–3100. [[CrossRef](#)]
10. Zhuge, Z.; Li, C. Dynamic pathways in the development of university students' AI literacy: Integrating quantitative and qualitative evidence. *Front. Psychol.* **2026**, *17*, 1822027. [[CrossRef](#)] [[PubMed](#)]
11. Liu, M.; Zhang, L.J.; Zhang, D. Enhancing student GAI literacy in digital multimodal composing through development and validation of a scale. *Comput. Hum. Behav.* **2025**, *166*, 108569. [[CrossRef](#)]
12. Zhong, B.; Liu, X. Evaluating AI literacy of secondary students: Framework and scale development. *Comput. Educ.* **2025**, *227*, 105230. [[CrossRef](#)]
13. Lintner, T. A systematic review of AI literacy scales. *npj Sci. Learn.* **2024**, *9*, 50. [[CrossRef](#)] [[PubMed](#)]
14. Hornberger, M.; Bewersdorff, A.; Nerdel, C. What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100165. [[CrossRef](#)]
15. Jaggy, A.-K.; Jansen, T.; Fütterer, T.; Trautwein, U. Almighty ChatGPT? Development of gifted secondary school students' AI (mis)conceptions. *J. Educ. Psychol.* **2026**, Advance online publication. [[CrossRef](#)]
16. Connell Pensky, A.E.; Eckstein, L.E.; Melville, M.C.; Pottmeyer, L.O.; Mineroff, Z.; Chawla, A.; Brooks, J.; Hershock, C.; Lovett, M.C. Can asynchronous learning modules about generative AI increase students' knowledge, skills, and self-efficacy for genAI? *Comput. Educ.* **2026**, *250*, 105640. [[CrossRef](#)]
17. Wang, Y.-Y.; Wang, Y.-S. Development and validation of an artificial intelligence anxiety scale: An initial application in predicting motivated learning behavior. *Interact. Learn. Environ.* **2022**, *30*, 619–634. [[CrossRef](#)]
18. Qiu, S.S.; Zhang, L.; You, F.; Zhao, X. Unpacking media channel effects on AI perception: A network analysis of AI information exposure across channels, overload, literacy, and anxiety among Chinese users. *Comput. Hum. Behav.* **2025**, *173*, 108790. [[CrossRef](#)]
19. Wang, Z.; Li, T.; Lyons, I.; Spangler, D.; Barroso, C.; Brockmole, J.R.; Quintero, M.; Wang, M.; Chaudhary, A.; Mazzocco, M.M.; et al. Emotion regulation and motivation modulate the associations between math anxiety, attention, and math performance. *J. Educ. Psychol.* **2026**, Advance online publication. [[CrossRef](#)]
20. Putwain, D.W.; Nicholson, L.J.; Pekrun, R. Test anxiety, emotion regulation, and achievement: Lagged and contemporaneous reciprocal relations. *J. Educ. Psychol.* **2026**, *118*, 240–258. [[CrossRef](#)]
21. Taboada Barber, A.; Cartwright, K.B.; Hancock, G.R.; Altani, A. Reading anxiety and reading achievement: Disentangling the role of reading engagement in the late elementary grades. *J. Educ. Psychol.* **2026**, Advance online publication. [[CrossRef](#)]
22. Cengiz, S.; Peker, A. Generative artificial intelligence acceptance and artificial intelligence anxiety among university students: The sequential mediating role of attitudes toward artificial intelligence and literacy. *Curr. Psychol.* **2025**, *44*, 7991–8000. [[CrossRef](#)]
23. Xu, T.; Wu, K.; Wang, D.; Wang, W. The double-edged sword of AI literacy: Functional differentiation of anxiety and motivational regulation in the GenAI context. *Learn. Motiv.* **2026**, *94*, 102276. [[CrossRef](#)]
24. Hsu, T.-C.; Chang, C.; Lin, Y.-W. Effects of voice assistant creation using different learning approaches on performance of computational thinking. *Comput. Educ.* **2023**, *192*, 104657. [[CrossRef](#)]
25. Song, X.; Zhang, Y.; Lu, Z.; Xu, L.; Shen, H. Generative AI: A double-edged sword for creative thinking learning—Evidence from facial expressions and fNIRS. *Comput. Educ.* **2026**, *247*, 105578. [[CrossRef](#)]
26. Ouyang, D. Voice, vulnerability, and expressive growth: Investigating AI anxiety and performance appraisal in voice arts education. *Front. Psychol.* **2025**, *16*, 1653502. [[CrossRef](#)] [[PubMed](#)]
27. Wei, L.-W. "I am creative, therefore I am employable": Self-creativity identity as a mediator of the AI-unemployment anxiety link. *Beijing Int. Rev. Educ.* **2025**, *7*, 224–238. [[CrossRef](#)]
28. Wu, X.; Tran, T.T.M.; Zhou, Q.; Sun, J.; Wu, H.; Wang, X. Human-AI co-creation or conflict? Mapping art students' diverse perspectives on creative identity with GenAI: A Q methodology study. *Educ. Inf. Technol.* **2026**, *31*, 2837–2876. [[CrossRef](#)]
29. Jou, M.; Hao, Y. Fostering career awareness and self-directed learning through an AI-Induced Disruptive Learning Environment. *Comput. Educ.* **2026**, *254*, 105688. [[CrossRef](#)]

30. Hou, C.; Zhu, G.; Sudarshan, V.; Lim, F.S.; Ong, Y.S. Measuring undergraduate students' reliance on Generative AI during problem-solving: Scale development and validation. *Comput. Educ.* **2025**, *234*, 105329. [[CrossRef](#)]
31. Zhuang, Z.; Li, X. The influence of collaborative music creation supported by generative artificial intelligence on students' creativity. *Front. Psychol.* **2026**, *16*, 1709513. [[CrossRef](#)] [[PubMed](#)]
32. Bewersdorff, A.; Hornberger, M.; Nerdel, C.; Schiff, D.S. AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students' AI self-efficacy. *Comput. Educ. Artif. Intell.* **2025**, *8*, 100340. [[CrossRef](#)]
33. Jiang, C.; Hui, B.; Sun, W.; Li, J.; Wang, X.; Wang, R. The psychological mechanism and heterogeneity of art design students' intention to use AI: An empirical analysis based on an integrated framework. *Front. Psychol.* **2026**, *17*, 1835335. [[CrossRef](#)] [[PubMed](#)]
34. Li, K.; Zhang, Y.; Ji, W. Is employment anxiety among vocal music students associated with AI replacement concerns? The roles of AI anxiety and vocal-performance replacement perception. *Front. Psychol.* **2026**, *17*, 1923159. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.