OPEN ACCESS
Journal of Advances in Education and Philosophy
Abbreviated Key Title: J Adv Educ Philos
ISSN 2523-2665 (Print) |ISSN 2523-2223 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Review Article

Integrating Intelligent Multimodal Technologies into Music Classroom Instruction: Application Scenarios, Pedagogical Pitfalls, and Improvement Pathways

Shuailin Chen[1]*

[1]School of Music and Dance, Xinyang Normal University, Xinyang 464000, China

DOI: https://doi.org/10.36348/jaep.2026.v10i06.001 | **Received:** 12.04.2026 | **Accepted:** 03.06.2026 | **Published:** 09.06.2026

***Corresponding author:** Shuailin Chen
School of Music and Dance, Xinyang Normal University, Xinyang 464000, China

Abstract

Addressing the inherent tension between the usability and instructional effectiveness of intelligent multimodal technologies in basic education music classrooms, this review focuses on core music learning activities in compulsory education and adopts a theoretical analysis approach. It constructs a systematic analytical framework consisting of "technological attributes, classroom scenarios, pedagogical pitfalls, and improvement pathways." The review argues that the value of intelligent technologies in music education does not lie in intensifying technology use, but in establishing structural alignment between technological support and authentic music tasks such as listening, performing, creating, and assessment. To address five prominent pedagogical pitfalls — goal displacement, resource overload, teacher capacity mismatch, suspension of learner agency, and insufficient assessment evidence — this review proposes five corresponding improvement pathways: a competence-based approach, scenario alignment, teacher guidance, activity reconstruction, and evidence construction. The theoretical contribution of this review is to integrate core competencies in music, classroom task design, teachers' professional judgment, and learning evidence construction into a unified analytical logic, thereby providing a conceptual reference for the digital transformation of music classrooms and for subsequent empirical research.

Keywords: Intelligent Multimodal Technologies, Music Classroom Instruction, Music Education, Competence-Based Learning, AI-Assisted Music Education.

INTRODUCTION

The rapid development of artificial intelligence, generative content production, learning analytics, and immersive media is driving school classrooms from the consumption of digital resources toward the construction of intelligent learning environments. As a typical multimodal learning field, the music classroom naturally involves the integrated processing of auditory perception, visual representation, bodily movement, linguistic expression, and cultural context. Consequently, when intelligent multimodal technologies enter music instruction, they create substantial opportunities for pedagogical enhancement while also generating deep tensions in instructional logic. This review focuses on music instruction within compulsory education, specifically at the upper primary and lower secondary levels, where advanced technological applications such as AI-assisted creation and digital audio editing are most prevalent. From the perspectives of the epistemological core of music and classroom pedagogy, it reexamines the boundaries within which technological tools can be integrated into music instruction.

Current classroom practice shows that traditional music instruction has long been constrained by limited resource formats, decontextualized listening activities, delayed feedback on performance, and insufficient conditions for creative work. Through the integration of cross-modal data, intelligent multimodal technologies provide technical leverage for constructing musical contexts, visualizing musical elements, enabling real-time assessment, and supporting digital creation. However, as large language models expand the possibilities of personalized support while also creating risks of cognitive dependence and outcome substitution (Kasneci *et al.,* 2023), high-density technological intervention in music classrooms may likewise allow technological spectacle to encroach upon aesthetic experience and the understanding of music itself.

Existing theories of multimedia learning and cognitive science provide a conceptual basis for analyzing this tension. Multimodal learning theory and cognitive load theory jointly indicate that although multi-channel information input can support knowledge representation, unstructured resource accumulation without clear goals and systematic selection may hinder learners' active processing under conditions of limited cognitive capacity (Mayer, 2024; Moreno & Mayer, 2007; Sweller *et al.,* 2019). This means that the pedagogical value of intelligent multimodal technologies depends fundamentally on whether they can build ordered cognitive scaffolds around melody, rhythm, timbre, and cultural meaning. Further research on music education technology confirms that the effectiveness of digital tools must be anchored in specific disciplinary activities in music. Whether in the construction of music TPACK frameworks (Bauer, 2013, 2020) or in the evaluation of technology-enhanced creativity (Lam, 2024; Liu *et al.,* 2025; Sanchez-Jara *et al.,* 2024), existing studies suggest that intelligent technologies can be transformed from external technological stimuli into internalized musical competencies only when they are embedded in authentic tasks such as listening, performing, creating, and responding, and when they are supported by effective feedback and learning analytics (Hamilton *et al.,* 2021; Hattie & Timperley, 2007; Siemens, 2013).

In summary, existing scholarship has established theoretical foundations for the present review from the perspectives of multimedia cognition, music pedagogy, and learning sciences. Nevertheless, relevant studies remain scattered across general educational technology and learning science research, while a systematic analytical framework oriented toward music classrooms in basic education is still insufficiently developed. At the same time, much of the existing literature emphasizes the theoretical potential of technology, but pays inadequate attention to pedagogical pitfalls such as goal displacement, suspension of learner agency, and insufficient assessment evidence in classroom practice.

This review extends prior work in three respects. First, whereas multimedia learning research explains how learners process multimodal information, this review situates that cognitive logic within the epistemological core of music, where sound, movement, notation, cultural context, and aesthetic response must be pedagogically coordinated. Second, whereas music TPACK research foregrounds teachers' integrated knowledge for technology use, this review further specifies how such knowledge should operate across concrete classroom scenarios, including resource generation, situated context construction, interactive inquiry, performance training, creative practice, and process-oriented assessment. Third, whereas digital pedagogy studies often emphasize technological affordances, this review foregrounds the pedagogical constraints and pitfalls that determine whether intelligent multimodal tools support or distort authentic music learning. Its added value therefore lies in shifting the analytical focus from tool adoption to task-technology alignment in music classroom instruction.

Based on this gap, the present review moves beyond the simple binary question of whether technology is effective and turns instead to the mechanism-oriented question of how technology becomes effective. Using theoretical analysis, this review clarifies the core application scenarios of intelligent multimodal technologies in music classrooms, diagnoses their pedagogical pitfalls, and proposes systematic improvement pathways from five dimensions: a competence-based approach, scenario alignment, teacher guidance, activity reconstruction, and evidence construction. To guide this systematic investigation, Figure 1 delineates the overarching conceptual framework, synthesizing the alignment mechanisms among technological affordances, musical disciplinary core tasks, pedagogical pitfalls, and improvement pathways. The review aims to provide theoretical reference and practical guidance for the digital transformation of frontline music teaching.

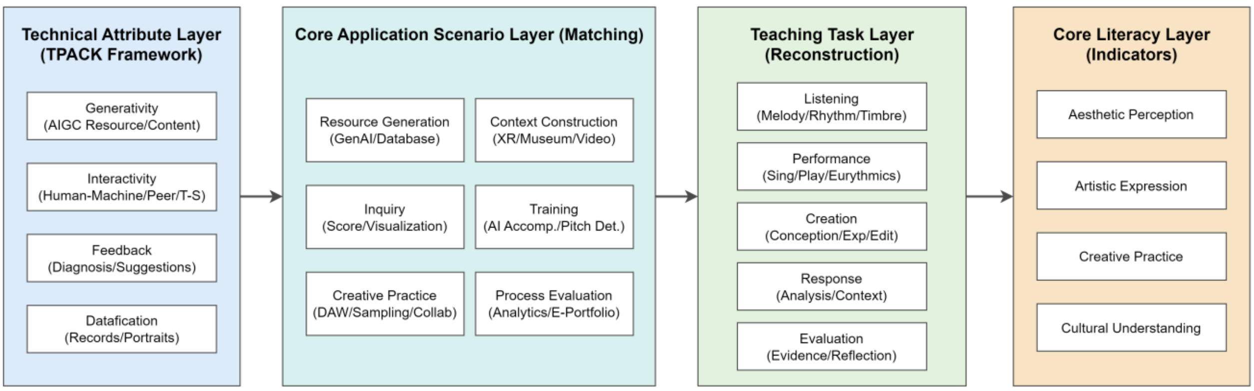

Figure 1: Conceptual Framework of Integrating Intelligent Multimodal Technologies into Music Classrooms: A Mechanism-Oriented Analysis

1. Conceptual Scope, Theoretical Basis, and Educational Value of Integrating Intelligent Multimodal Technologies into Music Classrooms

1.1 Conceptual scope of intelligent multimodal technologies.

Intelligent multimodal technologies refer to integrated technological systems that combine text, audio, images, video, movement, and interaction data while supporting content generation, real-time feedback, process tracking, and learning analytics. Compared with traditional multimedia technologies, their defining difference does not lie in the simple accumulation of media forms, but in their prominent features of generativity, interactivity, feedback capacity, and datafication. Traditional multimedia technologies are often limited by a one-way logic of presentation, demonstration, and explanation, leaving students largely in a passive receiving position. By contrast, intelligent multimodal technologies rely on generative AI, intelligent audio recognition, multidimensional visual representation, and learning analytics engines to reposition learners within an embodied learning cycle of inquiry, performance, creation, and assessment.

From the perspective of classroom pedagogy, these technologies are not value-neutral classroom "decorations." They are instructional support systems that must remain subject to curriculum goals, musical tasks, and teachers' professional judgment. Accordingly, the intelligent multimodal technologies discussed in this review go beyond individual tools. They refer to an intelligent learning ecosystem capable of comprehensively supporting resource generation, situated context construction, interactive inquiry, performance training, creative practice, and process-oriented assessment in music classrooms. To further clarify these distinctions, Table 1 provides a comparative analysis of the technological logic, information forms, and pedagogical roles associated with traditional multimedia versus intelligent multimodal technologies.

Table 1: Comparison between traditional multimedia technologies and intelligent multimodal technologies

Dimension	Traditional Multimedia Technologies	Intelligent Multimodal Technologies
Technological logic	One-way presentation and assisted explanation	Content generation, multidimensional interaction, real-time feedback, and data analysis
Information form	Linear combination of text, images, audio, and video	Deep integration of text, audio-visual materials, movement, and interaction data
Student role	Knowledge recipient and single-mode listener	Meaning inquirer, artistic performer, and creative practitioner
Teacher task	Courseware production and resource presentation	Task design, situated guidance, work diagnosis, and data interpretation
Core aim	Media presentation and background explanation	Construction of musical experience and development of disciplinary competencies

Cognitive science provides theoretical support for these features. Multimedia learning theory indicates that cross-channel information processing can promote understanding, but because cognitive capacity is limited, learning outcomes depend on whether learners can select, organize, and integrate information around core goals (Mayer, 2024). Moreno and Mayer further emphasize that interactive multimodal environments require appropriate guidance and feedback to reach deeper cognition (Moreno & Mayer, 2007). Cognitive load theory also warns that irrelevant information overload can seriously dissipate attention (Sweller *et al.,* 2019). These insights show that multimodal resources in music classrooms are not "the more, the better." If audio, video, scrolling comments, interactive displays, and AI-generated text are piled up without structure, students' attention can easily shift from listening to music itself toward consuming media stimuli, thereby weakening aesthetic depth. Therefore, the application of intelligent multimodal technologies must follow the principle of structured and restrained configuration around core musical elements (Macrides & Angeli, 2020).

1.3 Educational value of integrating intelligent multimodal technologies into music classrooms.

The educational affordances of intelligent multimodal technologies in music classrooms are mainly reflected in four dimensions.

First, they make implicit musical elements visible and concrete. In traditional teaching, voice relationships and formal structures are often too abstract, and students easily fall into generalized affective expressions such as "pleasant" or "cheerful." Through audio separation, rhythm visualization, and interactive maps, technology transforms transient sound into a learning object that can be retained and analyzed, providing media anchors for deeper comparison and aesthetic communication.

Second, they enable real-time feedback and precise intervention in performance training. According to feedback theory (Hattie & Timperley, 2007), effective performance training depends on the continuous narrowing of the gap between learning goals and current performance. Intelligent accompaniment, pitch detection, and recording playback overcome the limitation of single-point teacher feedback in traditional classrooms and provide students with frequent basic diagnosis. Teachers can then translate technological data into specific professional interventions, such as guidance on breath support and articulation.

Third, they expand the practical boundaries and expressive space of musical creation. With digital audio workstations and AI-assisted collaborative creation platforms, the technical threshold of complex arrangement is greatly reduced (Bauer, 2020; Lam, 2024). Students can complete an integrated experience of sound sampling, rhythmic creation, and multimedia production within limited classroom time, allowing technology to serve individualized creative expression rather than mere outcome generation.

Finally, they reshape the multidimensional chain of classroom assessment evidence. The intervention of learning analytics (Liu *et al.,* 2025; Siemens, 2013) extends assessment from single final performances to process data such as practice frequency, version iteration, and peer assessment. This provides data support for constructing a multidimensional assessment system oriented toward aesthetic perception, artistic performance, and cultural understanding.

Based on the educational value and multimodal cognitive logic discussed above, this review establishes a systematic analytical framework consisting of technological attributes, classroom scenarios, pedagogical pitfalls, and improvement pathways. The following sections use this framework to analyze core scenarios in which intelligent technologies enter music classrooms, the practical constraints they encounter, and corresponding improvement strategies.

2. Main Application Scenarios of Intelligent Multimodal Technologies in Music Classrooms

2.1 Typology of Technological Integration Scenarios

Resource generation: From teacher preparation to differentiated learning material design.

Resource generation is a preparatory scenario for the integration of intelligent technologies into teaching. Compared with traditional online searching, generative AI and multimodal content-processing tools can efficiently generate multiple versions of background maps, guiding questions, and tiered task sheets based on grade-level learning conditions, instructional goals, and musical genres, thereby providing support for differentiated learning. However, the educational value of this scenario does not depend on generation speed or material quantity, but on whether teachers can complete content curation, contextual calibration, and pedagogical translation. While large language models support personalized explanation, they may also produce misinformation, generalized content, and excessive dependence (Kasneci *et al.,* 2023). Therefore, AI-generated content should be positioned as a first draft for lesson preparation rather than as the final form of classroom materials. In listening or creative activities, teachers need to verify facts, revise language, and reconstruct tasks around musical genre, musical elements, and students' experience. The core issue in this scenario is not to expand the scale of resources, but to transform complex generated content into learning materials aligned with disciplinary goals in music.

Situated context construction: From visual rendering to cultural decoding.

Musical works are deeply rooted in specific historical and cultural contexts, and audio playback alone is unlikely to support students in constructing cultural meaning. XR, interactive video, and digital museum resources can reconstruct performance settings and cultural symbols, thereby providing immersive background for musical understanding. The classroom value of situated context construction lies not in the intensity of visual rendering, but in whether it directs students' attention back to the musical work itself. The instructional effectiveness of immersive technologies depends heavily on task design and learning measurement (Hamilton *et al.,* 2021). Without clear problem guidance, visual materials can easily displace musical listening. In practice, situated context construction should follow a progressive logic of "auditory primacy, contextual enrichment, structural analysis, and aesthetic expression." After multimodal materials enrich the context, students' attention should quickly be redirected to the analysis of melody, timbre, and formal features (Macrides & Angeli, 2020). Only in this way can context construction serve the holistic decoding of musical meaning rather than remain at the level of visual consumption.

Interactive inquiry: From passive appreciation to problem-driven analysis.

Traditional music appreciation teaching often faces the problem that students lack expressive vocabulary and their analysis remains at the level of vague impressions. With interactive scores, audio track separation, and rhythm visualization, abstract musical elements can be transformed into observable and comparable cognitive objects. In this scenario, interactive inquiry does not require students to click interfaces frequently or complete mechanical operations. Instead, it organizes their musical thinking through clear goals and progressively structured question chains (Mayer, 2024; Moreno & Mayer, 2007). Teachers may design inquiry tasks according to the logical sequence of "perception, analysis, expression, and culture": students first identify thematic motives and timbral features, then examine how tempo and mode shape musical imagery, and finally articulate their understanding through language, bodily movement, and cultural response. In this process, technological operation is embedded into core music response tasks, and students move from passive knowledge reception toward active interpretation of musical texts.

Performance training: From standardized demonstration to real-time feedback.

In singing and instrumental performance training, traditional teacher guidance is constrained by class size and often cannot provide comprehensive and frequent individualized intervention. Intelligent accompaniment following, pitch and rhythm detection, and video playback create a feedback cycle in which students can compare learning goals with current performance in real time. Effective feedback should not merely present gaps; it should also indicate the next step for improvement (Hattie & Timperley, 2007). If an intelligent platform provides only "accuracy rates" or "scores," students may easily fall into result-oriented thinking and neglect breath, timbre, and emotional expression in music. Therefore, performance training must maintain a dual-track mechanism of "objective machine feedback plus professional teacher diagnosis." Teachers need to translate platform-based quantitative data into concrete adjustment strategies, such as breath support, articulation position, and stylistic treatment, so that technological feedback can support the continuous refinement of artistic performance (Sanchez-Jara *et al.,* 2024).

Creative practice: From imitative reproduction to digital expression.

Digital audio workstations, AI-assisted arrangement, and collaborative creation platforms substantially lower the technical threshold of traditional music creation, enabling students in compulsory education to complete sound sampling, rhythm recombination, and multimedia production within limited time (Bauer, 2020). However, the educational value of creative practice is not equivalent to the complexity of the final product. It depends on whether students experience the process of conception, trial and error, revision, and explanation. Digital tools support creativity only when task design and assessment scaffolds work together (Lam, 2024). Therefore, the classroom must guard against AI generation replacing students' authentic thinking processes. Instructional design should preserve students' "thinking traces" by requiring them to submit creation statements, source records, and version iteration logs. For AI-generated musical fragments, students should also explain their purposes of use and revision intentions (Sanchez-Jara *et al.,* 2024). In this way, technology can function as a medium that expands the boundaries of sonic expression rather than as a "ghostwriter" that oversteps students' conception, experimentation, and reflection.

Process-oriented assessment: From summative judgment to evidence construction.

Intelligent technologies promote a shift in music assessment from summative judgment to process-oriented evidence construction. E-portfolios, practice trajectory tracking, and classroom interaction data can reduce the contingency of one-time performance assessment and record students' development across different stages. The meaning of learning analytics lies in explaining and optimizing the learning process, not in directly converting platform records into competency levels (Siemens, 2013). Because music is inherently performative and aesthetic, click frequency, practice duration, and automated scores can provide only limited evidence (Liu *et al.,* 2025). A more rigorous approach is to combine platform-based quantitative data with teachers' qualitative observations and students' self- and peer assessment to form a multidimensional evidence chain. Only by establishing a clear framework for interpreting assessment evidence and guiding students to understand the learning patterns behind data can intelligent multimodal technologies truly move music assessment toward performative and interpretive forms.

2.2 Task Alignment and Constraint Mechanisms

The six scenarios above show that the classroom application of intelligent technologies must follow the principle of task-driven integration rather than slide into technological instrumentalism. Different technological tools have specific functional boundaries. Only when they are deeply aligned with instructional goals and cognitive principles can they provide substantive support.

Table 2 functions as an analytical bridge between application scenarios and the subsequent discussion of pedagogical pitfalls. By aligning each scenario with its main teaching tasks and key constraints, it shows how technological affordances can become pedagogically productive only when bounded by disciplinary purposes. The five pitfalls discussed below are therefore derived from mismatches within this table: misalignment between technology and curriculum goals, excessive accumulation of resources, insufficient teacher mediation, replacement of learner agency, and weak assessment evidence.

Table 2: Correspondence among application scenarios, technologies, and teaching tasks

Application scenario	Suitable technologies	Main teaching tasks	Key constraints
Resource generation	Generative AI, digital resource platforms, multimodal repositories	Generate guiding questions, background materials, tiered exercises, and creative prompts	Teacher review of content accuracy, cultural appropriateness, and value orientation
Situated context construction	XR, digital museums, interactive video, high-definition performance resources	Establish work contexts, understand instrumental timbre, and connect regional culture	Avoid reducing instruction to sensory stimulation; serve understanding of the work
Interactive inquiry	Interactive scores, audio track separation, rhythm visualization, classroom interaction platforms	Analyze melody, rhythm, timbre, dynamics, and formal structure	Prevent technological operation from replacing musical analysis
Performance training	Intelligent accompaniment, pitch recognition, rhythm detection, recording playback	Improve singing, instrumental performance, rhythm, and movement	Platform feedback must be translated into concrete practice suggestions
Creative practice	Digital audio tools, AI accompaniment, sound sampling, collaborative creation platforms	Complete rhythm creation, melodic adaptation, scoring, and integrated works	Preserve students' conception, revision, and expressive processes
Process-oriented assessment	Learning analytics, e-portfolios, work version records, assessment rubrics	Construct evidence from practice, interaction, works, and reflection	Platform data cannot be equated directly with musical competence

In practice, these six scenarios are often combined within typical instructional tasks. For example, in folk song appreciation, teachers may first use generative AI to construct a basic cultural map, then use interactive scores to mark melodic contours and guide students to compare timbral and emotional treatment across versions. The final assessment must return to students' integrated interpretation of style and aesthetic elements. In rhythm training and sound sampling creation, intelligent feedback and digital arrangement tools should also be accompanied by explicit revision and reflection tasks, so that students do not drift into unreflective platform operation or simple score comparison. For lower primary students, complex tool operation should remain primarily on the teacher side, while students focus on listening, movement, and basic imitation, ensuring that technology use corresponds to learners' age-related cognitive characteristics.

Overall, intelligent multimodal technologies reshape the instructional ecosystem of pre-class preparation, in-class inquiry, and post-class assessment. Their core value does not lie in increasing technological density, but in using appropriate digital media to make implicit elements concrete, refine performance training, and make creative trajectories transparent. Only when technologies are tightly embedded in authentic musical activities and constrained by teachers' professional regulation can they avoid goal displacement and resource overload, thereby supporting the advancement of students' musical understanding and performance abilities.

3. Pedagogical Pitfalls in Integrating Intelligent Multimodal Technologies into Music Classrooms

3.1 Goal displacement

Technological display overrides musical understanding.

Once intelligent multimodal technologies enter the classroom, they can easily create a situation in which technological goals replace curriculum goals. Some teaching practices directly equate AI-generated content, interactive interfaces, or immersive images with classroom innovation, thereby weakening in-depth analysis of musical elements, formal structure, and cultural meaning. The fundamental aim of the music classroom is to cultivate competencies such as aesthetic perception, artistic performance, and cultural understanding, not merely to complete technological displays. If the instructional focus shifts toward tool operation, students may gain a brief sense of sensory novelty but fail to construct stable musical understanding. Digital pedagogy can generate real instructional value only when it is embedded in core musical activities such as creating, performing, and responding (Bauer, 2020). Once it is detached from music learning tasks themselves, the misconception that "the more complex the technology, the more advanced the classroom" leads to goal displacement.

3.2 Resource overload

Multimodal materials trigger cognitive load.

Although multimodal technologies provide abundant text, images, audio, video, and interaction data, unstructured resource accumulation does not necessarily improve teaching quality. From the perspectives of multimedia learning and cognitive load, learners can process information through visual and auditory channels, but their cognitive capacity is limited; the simultaneous presentation of excessive irrelevant information interferes with the processing of core content (Mayer, 2024; Sweller *et al.,* 2019). In music classrooms, if videos, scrolling subtitles, dynamic scores, AI question-answering, and real-time platform feedback appear simultaneously, complex media stimuli compete for students' attention. Music learning depends heavily on sustained attention to sound itself. If multimodal materials dominate the classroom, students' attention may shift from identifying acoustic details such as melody, rhythm, and timbre toward passive consumption of visual images, thereby weakening the depth of music learning.

3.3 Lagging teacher capacity

A disconnect emerges between digital literacy and disciplinary pedagogy.

Intelligent multimodal classrooms place higher demands on the integrated professional capacities of music teachers. Teachers' capacity for technology integration is not a simple transfer of basic digital skills; rather, it requires the synergistic application of technological, pedagogical, and music content knowledge (Bauer, 2013). In practice, some teachers possess basic tool operation skills but lack the instructional design capacity to transform tools into specific music learning tasks, or they cannot accurately interpret the internal relationship between platform data and students' musical performance. In addition, when faced with complex AI-generated materials, teachers who lack professional judgment regarding music history, stylistic features, and tool limitations may easily allow inaccurate content to enter instruction. Thus, the introduction of intelligent technologies does not lower the threshold of teaching; rather, it requires teachers to move from "courseware presenters" to "activity designers and data interpreters."

3.4 Suspension of learner agency

AI generation overrides authentic cognitive processes.

Although generative AI can significantly lower the threshold for music creation, it also risks suspending students' cognitive agency. The core value of musical creation and performance lies in the complete process of conception, trial and error, revision, and reflection, not merely in producing a standardized outcome. If students rely excessively on AI to generate melodies, accompaniments, or work descriptions with one click, the classroom may appear to have efficiently achieved its creative goals while in reality short-circuiting students' authentic cognitive and musical development. Similarly, in intelligent accompaniment or automated scoring environments, students may become dependent on machine feedback and mistake the improvement of accuracy scores for the only standard, neglecting emotional tension, timbre control, and stylistic treatment in musical performance (Sanchez-Jara *et al.,* 2024). If technology use lacks reasonable task constraints, it can shift from a learning scaffold to a substitute for students' own musical experience.

3.5 Insufficient evidence

Platform data cannot directly represent musical competence.

Intelligent technologies provide process data such as click rates, practice duration, and automated scores for classroom assessment, but such data cannot be equated with core competencies in music. Learning analytics should serve the interpretation and optimization of learning processes rather than produce seemingly precise assessment conclusions (Siemens, 2013). Because music is inherently performative and aesthetic, platform-based activity data often reflect only surface-level learning behaviors and cannot adequately capture students' ability to analyze work structure, the subtlety of aesthetic expression, or the unique intentions behind creative generation. Some platform indicators focus excessively on easily quantifiable pitch and rhythm while struggling to identify higher-order competencies such as emotional resonance, collaboration, and cultural understanding (Liu *et al.,* 2025). Therefore, excessive reliance on automatically generated platform scores leads to a thin assessment evidence chain and distorted assessment conclusions.

4. Improvement Pathways for Integrating Intelligent Multimodal Technologies into Music Classrooms

Adopt a competence-based approach and establish curriculum boundaries for technology integration.

To address goal displacement, the primary principle is to establish the baseline of "disciplinary competence first, technological tools subordinate." The starting point for choosing technology should not be "what functions can the existing tool provide," but "what musical competencies does this lesson aim to cultivate." More specifically, technology use should be examined at three levels of goals: at the level of knowledge understanding, whether technology highlights key musical elements; at the level of ability and performance, whether it supports singing, instrumental performance, or creative practice; and at the level of competence development, whether it deepens students' aesthetic experience and cultural connection. Only when technological functions are precisely aligned with core disciplinary tasks such as listening, performing, creating, and responding does their use gain educational legitimacy (Bauer, 2020).

Optimize scenario design through the parsimonious configuration of resources.

To address the cognitive load caused by resource overload, instructional design should move from the disordered accumulation of "many technologies in one lesson" toward a parsimonious configuration of resources. The presentation of multimodal materials should strictly control quantity, sequence, and timing, ensuring that every type of resource points to the same core musical task (Sweller *et al.,* 2019). In different scenarios such as resource generation, situated context construction, and performance training, technological media must serve specific instructional goals: situated context construction should return to the interpretation of a work's meaning rather than merely create sensory stimulation; interactive inquiry should converge on the analysis of musical elements rather than expand meaningless operations. Embedding technology into the cognitive sequence of music learning is the key to preventing attention dissipation.

Strengthen teacher guidance and deepen music TPACK and data interpretation capacity.

To address lagging teacher capacity, teachers must be supported in moving from "tool operators" to "designers of music learning environments." Teacher training should move away from simple software skill instruction and toward the cultivation of technology transformation capacity based on specific music tasks. The focus of training should not be only whether teachers can use a tool, but whether they can judge the musical accuracy and stylistic appropriateness of AI-generated content, and whether they can translate objective platform feedback, such as rhythmic deviation and pitch fluctuation, into adjustment strategies that students can understand and implement, such as articulation standards and breath control. Only by strengthening teachers' professional guidance and diagnostic intervention can the instructional value of intelligent technologies be realized (Bauer, 2013).

Reconstruct learning activities and protect students' agency in creation and performance.

To prevent AI from usurping or overshadowing authentic cognitive processes, classroom activities must preserve and make visible students' authentic thinking trajectories. In performance training and creative practice, teachers may construct a complete activity sequence of "listening perception, problem inquiry, technological support, work expression, and reflective assessment." Especially in AI-assisted creation, students should not directly submit a machine-generated final product. Teachers should require students to submit their original conception, version iteration records, sources used, and personal expressive intentions. By transforming the "traces of technology use" into learning evidence that must be examined, technology can remain a medium for expanding creative boundaries rather than a tool that replaces students' own thinking (Lam, 2024).

Improve the assessment system and construct a multidimensional evidence chain.

To overcome the limitation of single platform data, assessment must return to the disciplinary basis of music and construct a multidimensional process-oriented evidence chain. Platform data can only serve as supplementary reference and cannot replace professional judgment based on performance quality, aesthetic expression, and cultural understanding. The assessment system should integrate five types of evidence: listening records to anchor aesthetic perception; singing recordings and rubrics to evaluate artistic performance; version records and creation statements to verify creative practice; reflective texts to reveal cultural understanding; and interaction frequency to check learning engagement. Through cross-validation among platform-based quantitative data, teacher observation, and peer assessment, the problem that platform data cannot directly represent deep musical competence can be effectively addressed (Hattie & Timperley, 2007; Siemens, 2013).

CONCLUSION

Amid the digital transformation of education, the integration of intelligent multimodal technologies into music classrooms has become an important trend. Focusing on music teaching in compulsory education, this review has systematically identified six core classroom scenarios of technology use: resource generation, situated context construction, interactive inquiry, performance training, creative practice, and process-oriented assessment. The review reiterates that the usability of intelligent technologies is not naturally equivalent to instructional effectiveness. Their effectiveness depends on the guidance of authentic music tasks, the constraints of teachers' professional judgment, and the continuous optimization of classroom activity structures.

Addressing pedagogical pitfalls such as goal displacement, resource overload, lagging teacher capacity, suspension of learner agency, and insufficient assessment evidence, this review outlines a systematic improvement framework from five dimensions: curriculum goals, scenario configuration, teacher empowerment, activity reconstruction, and assessment improvement. Its fundamental implication is that music as a discipline must remain central, so that intelligent multimodal technologies can shift from external hardware accumulation and technological display toward an embedded instructional environment that supports students' aesthetic perception, artistic performance, and cultural understanding.

As a theoretical analysis, this review has constructed a systematic application logic framework, but it has not yet conducted empirical testing of specific intervention effects. Future research should define experimental conditions and assessment indicators more clearly, and use classroom empirical data, long-term tracking, and quantitative evaluation models to further clarify the mechanisms through which different technological scenarios influence the development of students' musical competencies. Such work would provide stronger empirical support for the digital transformation of music classrooms in basic education.

REFERENCES

- Bauer, W. I. (2013). The acquisition of musical technological pedagogical and content knowledge. *Journal of Music Teacher Education,* 22(2), 51-64. https://doi.org/10.1177/1057083712457881
- Bauer, W. I. (2020). *Music Learning Today: Digital Pedagogy for Creating, Performing, and Responding to Music* (2nd ed.). Oxford University Press. https://doi.org/10.1093/oso/9780197503706.001.0001
- Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education,* 8, 1-32. https://doi.org/10.1007/s40692-020-00169-2
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research,* 77(1), 81-112. https://doi.org/10.3102/003465430298487
- Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences,* 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274
- Lam, C. K. (2024). Technology-enhanced creativity in K-12 music education: A scoping review. *International Journal of Music Education,* 42(4), 691-703. https://doi.org/10.1177/02557614231194073
- Liu, Y., Tran, V. C., Kiss, B., Oo, T. Z., Szabo, N., & Jozsa, K. (2025). The use and effectiveness of digital tools in elementary music education: A systematic review. *Music & Science,* 8. https://doi.org/10.1177/20592043251363338
- Macrides, E., & Angeli, C. (2020). Music cognition and affect in the design of technology-enhanced music lessons. *Frontiers in Education,* 5, Article 518209. https://doi.org/10.3389/feduc.2020.518209
- Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review,* 36, Article 8. https://doi.org/10.1007/s10648-023-09842-1
- Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review,* 19(3), 309-326. https://doi.org/10.1007/s10648-007-9047-2
- Sanchez-Jara, J. F. M., Gonzalez Gutierrez, S., Rodriguez, J. C., & Syroyid, B. S. (2024). Artificial intelligence-assisted music education: A critical synthesis of challenges and opportunities. *Education Sciences,* 14(11), Article 1171. https://doi.org/10.3390/educsci14111171
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist,* 57(10), 1380-1400. https://doi.org/10.1177/0002764213498851
- Sweller, J., van Merrienboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review,* 31, 261-292. https://doi.org/10.1007/s10648-019-09465-5