Research on the Innovation of Art Education Teaching Models from a Multimodal Perspective

Shuang Guo*

Beijing Children's Palace, Beijing, 100062, China
**Correspondence email: 13811542506@139.com*

Abstract: *In the context of rapid evolution of information media and accelerated interdisciplinary integration, multimodal theory provides a new cognitive framework and methodological support for innovating teaching models in art education. This theory emphasizes the collaborative operation of multiple semiotic modes, including language, images, sound, movement, and spatial layout. Through the complementarity of sensory channels and cross-modal mapping of information, it reshapes learners' perceptual processes and meaning-making pathways. Based on an analysis of the connotations of multimodal theory, the cognitive attributes of art education, and the characteristics of symbols, this study proposes a construction pathway for art education teaching models from a multimodal perspective. The pathway includes mechanisms for the multimodal integration of teaching content, the diversified coordination of teaching media and technology, and the innovative design of teaching interaction patterns. Furthermore, the study explores optimization strategies such as the dynamic updating and expansion of teaching resources, the multimodal orientation of teaching evaluation systems, and the iterative and adaptive enhancement of teaching models. The research aims to construct an art education teaching model that is simultaneously structured, open, and flexible through a dual promotion of theory and practice, thereby providing diversified support for learners' aesthetic experiences, creative thinking, and comprehensive expressive abilities.*

Keywords: *multimodal perspective; art education; teaching model; resource updating; evaluation system; adaptive enhancement*

Introduction

Art education, situated in an era where visual culture and digital technology are increasingly intertwined, faces a deep transformation from single-medium instruction to multimodal integration. Multimodal theory, with its cross-symbol and cross-sensory information integration characteristics, offers both theoretical justification and operational pathways for reconstructing the teaching structure and interactive methods in art education. Traditional art instruction often centers on visual symbols, which effectively convey form and technique but remain limited in stimulating multi-channel perception and expanding cross-media expression. The introduction of a multimodal perspective allows multiple modes—such as images, language, sound, movement, and spatial layout—to form a collaborative network, enhancing learners' sensory engagement and expanding the possibilities for artistic expression. This study aligns with the trends of diversified media ecology in art education and personalized learning approaches, while cognitively addressing learners' needs for immersive, interactive, and cross-cultural art experiences, demonstrating significant theoretical and practical value.

1. Theoretical Foundations of Art Education from a Multimodal Perspective

1.1 Connotations and Developmental Trajectory of Multimodal Theory

Multimodal theory originates from the intersection of semiotics and communication studies. Its core idea emphasizes the collaborative operation of multiple semiotic modes in the process of meaning-making, regarding resources such as language, images, sound, movement, and spatial layout as equal and interactive carriers of meaning. From this theoretical perspective, information transmission no longer relies on linear delivery through a single medium, but instead achieves complex knowledge construction through the complementarity and interweaving of different sensory channels. The mapping relationships between information across different modalities enhance learners'

perceptual depth and breadth of understanding, enabling the formation of a more three-dimensional cognitive representation in environments with multi-channel input. This theoretical structure underscores that no single modality can independently convey complex cultural and artistic content; it must rely on the dynamic integration of multiple semiotic forms to realize multidimensional meaning generation and deep construction [1].

In the field of education, the introduction of multimodal theory has promoted the transformation of teaching from single-medium instruction to diversified and interactive approaches. The theory not only alters the media structure of knowledge presentation but also reshapes learners' cognitive processing pathways and mechanisms of meaning generation. In art education, multimodal theory positions visual symbols no longer as the sole core; various forms of information, including verbal explanation, musical accompaniment, physical demonstration, and digital interaction, can form mutually complementary knowledge networks within a single teaching scenario. Such multimodal integration not only enriches learners' sensory engagement in art education but also provides diverse support for emotional mobilization and aesthetic judgment during the creative process, establishing a solid theoretical foundation for innovating art education teaching models.

1.2 Cognitive Attributes and Semiotic Features of Art Education

The cognitive structure of art education exhibits dual characteristics, in which sensibility and rationality operate in parallel, and intuition and logic complement each other. Sensible cognition directly engages visual and emotional symbols through elements such as color, form, and texture, stimulating learners' aesthetic experiences and creative impulses. Rational cognition, on the other hand, relies on systematic rules such as perspective, proportion, and composition to analyze and integrate perceptual materials, forming controllable artistic expressions. This dual-track cognitive mechanism renders art learning a dynamic process where sensory perception and logical reasoning intertwine. Learners are required not only to experience the immediate activation of imagery and emotion but also to achieve structured expression within the framework of a symbolic system, thereby unifying creation and comprehension through the interaction of perception and thought.

The semiotic system of art education demonstrates high dimensionality and transferability. Color can evoke multi-layered emotional resonance through parameters such as hue, saturation, and brightness; forms convey implicit narratives and order through proportion, rhythm, and symmetry; composition guides viewers' attention and thought flow via focal arrangement, visual pathways, and spatial segmentation. These symbols are not confined to the visual domain but can establish cross-modal connections with verbal narrative, musical rhythm, or bodily movement, generating a more expressive artistic context. Within a multimodal framework, this capacity for cross-domain integration of symbols greatly expands the mediums and boundaries of art education, transforming artistic creation from unidimensional visual presentation into a multi-channel, multi-layered comprehensive expressive system [2].

1.3 Theoretical Logic of Integrating Multimodality with Art Education

The integration of multimodality and art education is based on the dual principles of semiotic resource integration and the complementarity of sensory channels. Through a closed-loop structure of multimodal input and output, learners can mobilize multiple senses to decode information during perception and employ different modalities for re-encoding during creation. For example, the combination of images and language enriches the narrative layers of artworks; the emotional resonance between color and music strengthens the affective impact of art; the coordination of spatial layout and bodily movement creates an immersive aesthetic experience. This cross-modal collaborative mechanism enhances aesthetic understanding at the perceptual level while expanding diversity and innovation in expressive practices at the creative level.

From a theoretical perspective, the integration of multimodality with art education not only entails the updating of teaching methods but also represents a reconstruction of learners' cognitive structures. In a multimodal environment, semiotic resources do not exist in isolation; they form a highly interconnected whole through semantic cohesion, temporal synchronization, and functional complementarity. The process by which learners transfer meaning and generate innovation across different modalities is, in essence, a dynamic practice of multi-sensory integration and cross-domain thinking. This integration model emphasizes constructing teaching scenarios with layers and openness based on learners' perceptual characteristics and cognitive principles, enabling art education to achieve

simultaneous advancement in aesthetic guidance, skill development, and cultural interpretation, while providing a solid theoretical foundation for the continuous evolution of future teaching models.

2. Construction Pathways for Art Education Teaching Models from a Multimodal Perspective

2.1 Multimodal Integration Mechanism of Teaching Content

The multimodal integration mechanism of teaching content aims to achieve a three-dimensional presentation of knowledge structures and multi-layered transmission of informational meaning through the organic collaboration of different sensory channels and semiotic systems. In art education, this mechanism encompasses not only traditional visual elements such as color, line, and composition, but also verbal explanation, background music, dynamic images, physical demonstrations, and tactile experiences. By situating these elements within a unified teaching context, a multi-sensory resonant learning environment can be created, enabling learners to activate visual, auditory, and tactile channels simultaneously while acquiring information, thereby enhancing perceptual processing depth and the capacity for generating artistic language. This integration improves knowledge acquisition efficiency and enriches artistic understanding with cultural and emotional connotations.

The key to multimodal integration lies in structured collaboration between modalities rather than superficial overlay or collage. In instructional design, different modalities should achieve precise functional division and semantic cohesion based on knowledge attributes, teaching objectives, and cognitive principles. For example, the warm-cool variation of color can correspond emotionally with musical fluctuations; perspective and composition analysis can be synchronized with teachers' verbal explanations and gestural demonstrations; material texture presentation can be paired with tactile experiences to deepen memory and comprehension. By precisely controlling temporal rhythms and spatial arrangements, the content of different modalities can form coherent streams of meaning, allowing learners to develop three-dimensional cognition through multi-channel information interweaving and to cultivate higher-level artistic creativity through diversified aesthetic perception [3].

2.2 Diversified Coordination of Teaching Media and Technology

The construction of multimodal art education relies on the deep coordination of diversified media and technology. With the development of digital and intelligent technologies, the forms of instructional information and modes of interaction have become unprecedentedly diverse. The introduction of virtual reality (VR), augmented reality (AR), and three-dimensional modeling technologies enables learners to explore the details and spatial logic of artworks in immersive contexts. The combination of multimedia projection and interactive screens provides classrooms with adjustable, real-time visual presentation channels. Touch-sensitive and sensor-enabled devices not only expand learners' modes of interaction but also enhance engagement and creative freedom through motion feedback mechanisms. The integration of these technologies within a single instructional system achieves comprehensive mobilization of visual, auditory, and kinesthetic senses, providing a solid foundation for precise multimodal information transmission.

The coordination of diversified media extends beyond hardware and software compatibility to the optimization of information transmission pathways and instructional content organization. In art education, the integration of digital and traditional media can create cross-temporal and cross-spatial artistic experiences, such as combining high-definition digital reproductions of classic artworks with live painting activities, enabling learners to switch freely between virtual and real contexts and deepen understanding and creativity across dual settings. Multimodal technology design should balance sensory load and information fluency, avoiding cognitive interference caused by excessive technological elements. Only by maintaining dynamic tension between technological implementation and aesthetic objectives can teaching models achieve functional upgrading and expressive extension, establishing an efficient reciprocal relationship between media application and artistic experience in art education.

2.3 Innovative Design of Teaching Interaction Models

Within a multimodal perspective, the innovation of teaching interaction models emphasizes multi-channel and multi-layered communication and meaning co-construction between teachers and students as well as among students. Interaction in art classrooms extends beyond verbal questioning or

artwork presentation to networks of information exchange and collaborative creation constructed through image annotation, dynamic demonstration, musical feedback, and gesture capture. This interactive model strengthens learners' agency, enabling a cyclical cognitive processing loop in perception, analysis, and expression, thereby facilitating more natural and fluid creative generation. Moreover, multimodal interaction transcends spatial and temporal limitations, making remote collaboration, synchronous creation, and cross-cultural artistic exchange possible, thus expanding the spatial dimension and cultural reach of instruction [4].

The core of innovative interaction lies in the immediacy and diversity of feedback. For example, in a class analyzing an artwork, teachers can use a digital whiteboard to illustrate compositional analysis paths in real time while providing verbal explanation and gestural guidance, allowing students to follow and comprehend synchronously. Students can respond instantly through tablets or digital drawing boards on shared interfaces, creating or modifying content collaboratively. During cooperative creation, different learners can manage different modal information processing tasks, such as color coordination, musical arrangement, and spatial composition, ultimately integrating these elements into a comprehensive artwork with visual, auditory, and spatial impact. Through such a multimodal interactive framework, the teaching process evolves from unidirectional information transmission into a generative field of shared creativity, emotional resonance, and meaning reproduction, providing sustained impetus and potential for the continuous innovation of art education models.

3. Optimization Strategies for Art Education Teaching Models from a Multimodal Perspective

3.1 Dynamic Updating and Expansion of Teaching Resources

The continuous optimization of multimodal art education primarily relies on the dynamic iteration and systematic expansion of the teaching resource system across multiple dimensions, including quantity, quality, structure, and presentation. Unlike traditional static and closed resource repositories, updates aimed at multimodal expression involve not only the supplementation of materials but also comprehensive structural reorganization and functional reconstruction. In this process, various types of resources—such as static images, high-definition scans of artworks, dynamic videos, virtual 3D models, sound narratives, interactive animations, and immersive environmental materials—are recombined to form cross-sensory information pathways, enabling learners to obtain richer artistic experiences through integrated visual, auditory, tactile, and even olfactory perception. The core of dynamic updating lies in constructing an intelligent, semantically driven, and sustainably evolving resource management mechanism. Through tagging, modularization, and semantic retrieval, resources can be efficiently located and precisely deployed, ensuring high alignment with course themes, cognitive stages, and individual differences [5].

The expansion of resources extends beyond diversification of material sources to include deep integration of interdisciplinary and cross-cultural content. Artistic symbols from different cultural contexts, cutting-edge explorations in technology-based art, and forms that combine emerging fields such as artificial intelligence and data visualization can all inject new creative potential and visual language into the resource repository. For instance, integrating traditional painting and calligraphy with augmented reality (AR) technology can reconstruct the dynamic expressiveness of classical art in virtual space; combining digital sculpture with 3D printing allows students to experience the full cross-media creative process within the "virtual design–physical generation" workflow. By introducing cloud-based sharing platforms and geographically distributed material databases, resource updating can transcend temporal and spatial limitations, cultivating both a global perspective and localized sensibilities. Such dynamic updating and expansion not only enrich the semiotic system of art education but also establish a solid media foundation for future model innovation and knowledge structure reconstruction.

3.2 Multimodal Orientation of Teaching Evaluation Systems

Within a multimodal perspective, the traditional unidirectional evaluation model that relies solely on the quality of final works can no longer comprehensively reflect learners' artistic abilities and creative potential. The new multimodal evaluation system emphasizes the incorporation of full-process tracking and multi-channel data collection into the assessment framework, enabling synchronous analysis of multiple dimensions, including visual composition, color variation, spatial perception, auditory accompaniment, movement trajectories, and verbal narration. The objects of evaluation extend

beyond static outcomes to include the evolution of creative ideas, iteration of sketches, experimental records of media switching, dynamic adjustment of color schemes, and learners' immediate verbalized conceptions. This process not only reconstructs learners' cognitive pathways and thinking rhythms but also provides teachers with quantifiable, traceable, and fine-grained process data to optimize instruction.

The construction of this system requires balancing scientific rigor with artistic sensitivity between quantitative and qualitative approaches. Quantitative analysis can rely on digital assessment tools to precisely measure aspects such as color distribution frequency, compositional proportion changes, diversity and innovation of media usage. Qualitative evaluation focuses on elements such as emotional resonance, symbolic innovation, cultural significance, and thematic consistency of the works. By establishing an evaluation model based on multimodal data integration, teachers can generate personalized feedback reports and formulate targeted improvement suggestions for different types of learners. This orientation renders the evaluation system more multidimensional, dynamic, and flexible, while transforming the assessment process into a key driver for learners' self-reflection and creative strategy optimization, thereby holding significant value in promoting the integrated development of technical skills, creative thinking, and aesthetic judgment.

3.3 Iteration and Adaptive Enhancement of Teaching Models

The iteration of multimodal art education teaching models constitutes a systematic optimization process driven by data and feedback regulation, encompassing the coordinated evolution of multiple core elements, including instructional content design, media and technology configuration, interaction structure construction, and evaluation feedback mechanisms. Unlike the linear replacement characteristic of traditional models, this iteration emphasizes the optimization of informational presentation logic and the coordination of sensory stimulation through structural adjustments while retaining core educational objectives. For example, when learners demonstrate low responsiveness to a specific type of media, teachers can adjust the proportion of multimodal combinations; after introducing new technology, experimental courses can be used to verify its adaptability across student groups of different ages and cultural backgrounds. The sustainability of iteration relies on the openness and scalability of the model, ensuring that core structures remain stable while new media and technologies are integrated efficiently with existing resources and methods [6].

The core of adaptive enhancement lies in increasing the model's flexibility and agility to accommodate diverse learning contexts and individual differences. In response to variations in learners' perceptual preferences, technological familiarity, cultural backgrounds, and artistic style orientations, the teaching model can dynamically adjust modality selection, information density, and interaction pacing. For instance, for beginners, the complexity of parallel multimodal presentation can be reduced to emphasize guided experiences led by a single sensory channel; for learners with foundational skills, the complexity of cross-modal creative tasks can be increased to encourage media integration and innovative expression. By establishing a parameterized and configurable model structure, art education can achieve personalized adaptation while maintaining consistency in core educational values, enhancing inclusivity and precision in instruction, and providing a solid foundation for future cross-disciplinary artistic creation and the cultivation of versatile talents.

Conclusion

The innovation of art education teaching models from a multimodal perspective involves not only the integration of media and technology but also the reconstruction of cognitive structures and modes of expression. By establishing a dynamically updated teaching resource system, optimizing a multimodal evaluation mechanism encompassing the entire creative process, and constructing teaching models with iteration and adaptability, art education can achieve a shift from single-sensory-driven instruction to multi-channel collaborative engagement. This model enhances learners' perceptual depth and creativity while providing a flexible structural foundation for cross-disciplinary integration and international expansion in art education. Future development may further explore the deep integration of emerging technologies, such as artificial intelligence and immersive interaction, with multimodal art education, thereby constructing a more intelligent, personalized, and open learning ecosystem to promote the continuous evolution and value extension of art education within the digital cultural environment.

References

[1] Tian, X. "Research on the Innovation of Higher Vocational Art Education Teaching Models Based on 'Internet+'." Craftsman 05 (2025): 27-29.
[2] Wu, K., Jin, B., and Zhou, Q. Y. "Exploration of Media Innovation and Teaching Model Transformation in Art Education in the Era of Artificial Intelligence." Teacher Education Forum 37.11 (2024): 28-35.
[3] Liu, L., and Liu, Z. "Research on Innovative Paths of Color Teaching in University Art Education." Art and Era (China) 07 (2024): 102-104.
[4] Cheng, X. Y. "Application of Diversified Teaching Models in Art Courses for Early Childhood Education Majors." New Curriculum Research 12 (2023): 44-46.
[5] Shen, M. J. "On the Relationship Between Innovation and Entrepreneurship Education and Teaching Models in Art Majors." Science & Technology Information 20.15 (2022): 167-169.
[6] An, R. "Discussion on the Construction of Innovative Art Education Models in Universities." Art and Era (China) 05 (2021): 93-94.