

Chinese High School Teachers' Perceptions and Recommendations of Using AI-Assisted Visual Text Multimodal in Art Classes

Tiange Xiang, Khairul Azhar Bin Jamaludin*

Faculty of Education, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia

Email: p145816@siswa.ukm.edumy, khairuljamaludin@ukm.edu.my

DOI Link: <http://dx.doi.org/10.6007/IJARPED/v15-i1/27391>

Published Online: 09 January 2026

Abstract

The surge of artificial intelligence has led to plentiful opportunities to reform education in new ways which might not have happened in the past. Multimodal instructions become available through the generation of visual images and animations that can be best accompanied with instructional delivery and feedback, especially in arts classes where visuals are frequently emphasized. However, despite the effectiveness, understanding the art teachers' experiences of implementing these tools remain under-researched. A qualitative exploratory research design is utilized, specifically through a set of interview questions designed for a group of high school teachers in a selected educational institution in Zhejiang, China. Thematic analysis is used to present the findings, informing the perceptions and experiences of using AI, the challenges of using the tool, and the subsequent recommendations shared by these teachers. In terms of the significance of the study, it contributes to informing the state of artificial intelligence in high school art education contexts and highlighting how support can be given as a form of mediator to boost teachers' professionalism in teaching.

Keywords: Artificial Intelligence, Ai Assistance, Art-Based Learning, Visual Text, Visual Text Multimodal

Introduction

The recent years have witnessed the surge of artificial intelligence (AI) and its usage in various educational settings, effectively changing the landscape and reshaping how pedagogical methods are employed in the teaching and learning processes in schools. The designing of instructional materials, delivering the contents, and experiencing the learning processes are mainly affected by the introduction of AI. Particularly, the use of generative AI technologies, such as the likes of ChatGPT, Gemini, and CoPilot, have become one of the core instructional materials in art classes, as they greatly assist in visual learning for the generation of images, animations, and texts. AI, then, allows significant student engagement in classrooms, subsequently developing a conceptual understanding of the contents to be learned beyond utilizing the traditional way of textbooks and paper notes to acquire knowledge (Chen et al., 2024; Guan et al., 2025).

In the larger sense and beyond the context of the learning of arts, English language learning continues to remain a challenge for many students in China (Tang & Leong, 2025). Despite the ever-growing importance of acquiring and using English not only in classrooms but also in workplace settings, English remains as a foreign language in the Chinese setting. Although there are signs informing English is gradually becoming a second language in China as a result of changes in the educational policies of English language learning, continuous efforts are being made to deliver quality English lessons while resolving the challenges of English language acquisition in all levels of education in China (Tang & Leong, 2025). While AI is efficient in providing visual contents to assist students in acquiring arts-related knowledge, the same advanced technology is also capable of delivering plentiful language learning contents to the students.

Problem Statement

Notwithstanding the potential strengths offered by artificial intelligence in all levels of education, including arts education, implementing AI in classrooms remains a challenge to many teachers in China, especially in terms of its implementation and tailoring the content generated to the existing curriculum objectives and meeting students' educational needs and goals. Plentiful studies concerning arts education have been focusing on emphasizing student engagement, creativity, and learning goals and outcomes (Ng et al., 2025; Zhang, 2025). Consistent with the literature available of teaching using AI, many Chinese teachers continue to report the struggles they face when it comes to making AI purposeful for lessons and integrating it to achieve learning objectives, especially among the art education teachers when it comes to their experiences of interpreting AI and its integration into the classroom setting (He & Zhang, 2025; Tang & Leong, 2025).

Furthermore, another problem that arises is aligning the AI-generated contents with the learning goals in art education. Specifically, students and teachers are guided by the curriculum specifications and the corresponding subject and lesson objectives. Teachers are also expected to adhere to all curriculum standards outlined by the Ministry of Education, displaying the necessary pedagogy by integrating AI and other emerging technologies in the classroom setting to boost art education learning (Ng et al., 2025; Zhang, 2025). However, despite the rich content generation by AI, putting and ensuring these contents to match the specifications and objectives aforementioned can be challenging, especially when mismatches occur. These mismatches are also expanded when AI-generated contents fail to align themselves with the curricular goals, sequences of teaching and learning in a classroom, and the assessment requirements (Zhang, 2025). When mismatches occur, students may become confuse, failing to see the knowledge that they are supposed to acquire from the lessons they have been attending for long periods or within the term (Tang & Leong, 2025). These students may also fail to recognize how the contents learned can be useful in future settings, especially in the workplace (Alsohaimi et al., 2025; Tang & Leong, 2025). This posits a strong need to understand the exact experiences teachers have gone through, exploring how their lived experiences in their teaching career with AI in relation to the teaching of arts in the Chinese education setting (Zhang, 2025).

Arguably, existing research has been focusing on how AI can be geared towards achieving student outcomes in almost all levels of education over the recent years. Plentiful studies have revealed valuable insights of how AI can be used to support the learning of different

subjects across different levels of education. However, the perspectives of teachers who are delivering the subjects and the experiences of using AI remain under-investigated, especially at the secondary level and vocational level of arts education (Cohn et al., 2024; Tang & Leong, 2025). How these teachers accept and negotiate with the use of AI, alongside their reasoning with these emerging tools in their teaching career, must be investigated further to ensure the effective implementation of AI in arts education settings. The challenges that these teachers may face in classrooms must be explored further, as these experiences inform the restrictions and constraints they are currently facing so as to allow policy interventions and allocation of specific training sessions to raise the teachers' standards (Cohn et al., 2024; Tang & Leong, 2025).

Understanding the teachers' experiences of using AI to deliver art lessons in classroom is crucial in informing current experiences faced by these teachers to provide actionable, timely interventions and solutions to improve pedagogical capabilities and lesson delivery (Cohn et al., 2024). Understanding these teachers' experiences in classrooms is necessary to provide a clear depiction of the state of AI adoption and integration in art classrooms, thereby reducing the risks of hampering the progression of learning without sufficient support and encouragement from the institution and the government (Ng et al., 2025). Furthermore, the role of teachers in arts education is not just improving the students' creativity to expand their talents in future creative industries (Cohn et al., 2024; Tang & Leong, 2025; Zhang, 2025). They also need to ensure the students have the necessary English language capabilities to participate in the creative industries, not only in China but also in global contexts.

Significance of the Paper

Significantly, this paper provides substantial contributions from a theoretical, practical, and institutional lens. First, theoretically, this study informs how AI-assisted visual text multimodal is supported by the Multimodal Learning Theory to help teachers in delivering AI in arts education, thereby also informing the extent of teacher agency and their classroom control in the school setting. Then, from a practical aspect, the needs of the teachers can be addressed, highlighting the necessary interventions institutions must provide to assist with the professional development among these teachers and ensure that AI-generated contents are aligned with the curriculum standards and the specific lesson goals (He & Zhang, 2025; Tang & Leong, 2025; Zhang, 2025). Finally, in terms of institution, policymakers, school leaders, key authorities in the Ministry of education can understand the voices of these teachers, thereby striving to fulfil the teachers' needs and concerns while also achieving sustainable AI implementation and integration in contemporary classroom teaching and learning (Alsohaimi et al., 2025). At the same time, these parties will also be able to inform sound policy decisions that is sensitive to the classroom teaching and learning process, ensuring quality education is delivered (Alsohaimi et al., 2025).

Research Objectives

In light to the identified problems, two research objectives were established to guide the researchers in this paper. While demonstrating the competence of using English is crucial for the students' participation in future workforce, this study centralizes its focus on the teachers' experiences of using AI instead. First, the researchers explored the perceptions of high school teachers' perceptions of using AI-assisted visual text multimodal to teach art classes. Then, the researchers explored the same group of high school teachers' recommendations of using

the same tool to teach art classes. More specifically, for each objective, the researchers established the sub-research objectives that provided a clearer guidance to depict the state of using AI-assisted visual text multimodal in the teaching of art classes.

Main Research Objective 1: To explore the high school teachers' perceptions of using AI-assisted visual text multimodal to teach art classes.

Sub-Research Objective: To explore the challenges of using AI-assisted visual text multimodal to teach art classes.

Main Research Objective 2: To explore the high school teachers' recommendations of using AI-assisted visual text multimodal to teach art classes.

Sub-Research Objective: To explore the practical recommendations of using AI-assisted visual text multimodal to teach art classes.

AI-Assisted Visual Text Multimodal

The presence of artificial intelligence has led to the emergence of using chatbots to produce various educational resources and materials. As a multimodal, AI is successful in encouraging different modes of content creation and meeting different learning needs. With the assistance of AI, multiple data can be processed and content can be generated, allowing deeper understanding to occur for human comprehension (Christoforou, 2025; Tan et al., 2025). In the context of teaching and learning, the content generated facilitates meaningful student-centered learning by providing images, animations, and words for students to use (Christoforou, 2025; Tan et al., 2025). Students continue to interact with the stimuli during and after lessons, further receiving feedback and guidance from both the teachers and artificial intelligence itself (Christoforou, 2025; Tan et al., 2025).

For teachers and educators, and sometimes to the students, using AI-assisted visual text multimodal also encourages them to be creative and analytical when writing prompts to AI models during the process of generating contents (Christoforou, 2025; Tan et al., 2025). It requires a fusion of existing knowledge in arts and their intended objectives for the AI to generate contents. Any output produced would reflect the extent of knowledge the teachers and students possess.

Multimodal Learning Theory

Concerning the use of AI in the teaching and learning in the field of arts, the Multimodal Learning Theory (MMLT) best acted as a theoretical guidance in this study. Richard Mayer developed the Multimodal Learning Theory, asserting that learning experiences are best enhanced through the integration of different senses rather than relying on one (Kress & van Leeuwen, 2001; Mayer, 2009; Paivio, 1986; ; Vatrál et al., 2023). MMLT posits the use of multiple sensory modalities, including, but not limited to, visual, textual, and auditory senses to achieve effective learning. Furthermore, proponents of MMLT suggest that meaningful learning occurs when these senses are combined to deliver the lesson content during lessons (Al-Jarf, 2024; Arifin et al., 2025). Artificial intelligence thus conforms to the principles of MMLT in utilizing multiple senses to deliver lesson content, as AI can produce visual images and texts that are effective in supporting comprehension of ideas and creating lasting memories of the contents learned, which is arguably significantly better than merely looking at texts (Al-Jarf, 2024; Vatrál et al., 2023). Simply, the use of AI allows greater knowledge

retention instead of relying on the typical and traditional textbook-and-paper teaching method (Al-Jarf, 2024; Arifin et al., 2025; Vatral et al., 2023).

Within the context of art learning in China, the researchers believed that MMLT offers a dual coding process, wherein at least two codes, or in this study, the senses of visual and textual aspects, that allow students to acquire and construct knowledge by constantly interacting with these senses. In addition to the dual coding process, MMLT also posits learning as an engagement through kinesthetic and tactile actions, require actual hands-on performances and application of knowledge into creating works and completing assignments (Al-Jarf, 2024; Arifin et al., 2025; Vatral et al., 2023). However, in actual situations, using AI for learning purposes transcends the dual-coding process, as students not only get to be visual learners by observing images and texts generated by AI, but they are also able to be auditory learners due to the read-out-loud function available in AI (Arifin et al., 2025), as shown in AI chatbots like ChatGPT and Gemini. Evidently, the contents generated by AI can also be used for students to engage in problem solving and decision making. Through the assistance from AI, students are not merely passive receivers of knowledge, as careful designing and prompting from teachers would allow students to be active learners in the classroom.

Due to the generative nature of artificial intelligence, students can interact with the lesson contents constantly and apply their knowledge into the tasks provided by AI. Feedback occurs through the evaluation and assessment of the students' performance in AI itself, where AI provides accurate feedback on the mistakes made by the students, informs constructive feedback, and shares suggestions for improvement. Hence, accompanying the multiple senses available and the corresponding feedback given, learning of arts becomes increasingly accessible and effective for various students in a classroom setting, catering to their different learning needs.

Research Methodology

For the purpose of this study, the qualitative exploratory research design was selected and utilized. This design was necessary to aid the researchers in understanding how a certain phenomenon was experienced (Creswell & Guetterman, 2019; Hunter et al., 2019; Olawale et al., 2023); in this study, it refers to the use of AI-assisted visual text multimodal in the teaching and learning of art classes. By utilizing the qualitative exploratory research design, the subsequent qualitative inquiries employed would allow a rich and contextualized description of the data obtained, especially through inquiries like interviews, document analysis, and observations. Furthermore, one of the core implications of using this design is to also uncover potential new insights of a phenomenon or idea, allowing suggestions for policy changes and future research (Creswell & Guetterman, 2019; Edmonds & Kennedy, 2017).

Sample Selection

This study was conducted in Zhejiang Art Vocational College, which was located in the Zhejiang province in China. This college, whilst provided vocational education and tertiary education for students interested in the field of arts and designing, it also provided similar education to secondary school students, providing them with the opportunities to strengthen their art talents before entering the creative industry. Graduates from this college would have the opportunity, but not limited to, participate in art designing jobs, theater and film

production, and television performances related to art. With the emergence and continuous usage of artificial intelligence, the college welcomed its usage and implementation in its educational settings, thereby allowing the researchers to conduct the study in the college. For the purpose of this study, the researchers recruited five high school teachers from Zhejiang Art Vocational College. As later exemplified and explained, the pseudonym TOX was used to refer to the high school teachers. Generally, all five teachers had a major in arts, but they were also tasked to teach English lessons to the students.

Research Instruments

A set of semi-structured interview questions was used to achieve the corresponding research objectives. The interview questions were established and designed based on the literature available of using artificial intelligence in the art educational setting. The interviews were conducted in a face-to-face manner in the college, subjected to the high school teachers' convenience and availability for the interview. These teachers were informed with the research purposes, and their participation was voluntary. During the interviews, an interview protocol informing the steps to conduct the interview before, during, and after the interview, alongside a recorder that was used to record the session. Before the interviews, the high school teachers were also briefed with the use of the AI-assisted visual text multimodal and its corresponding usage.

To analyse the data, Braun and Clarke's six steps of thematic analysis were referred to and used. This qualitative data analysis method is widely used in majority of mixed-method and qualitative research (Ahmed et al., 2025; Maguire & Delahunt, 2017). In this method, the six steps are (1) familiarizing with the data, (2) generating initial codes, (3) grouping codes to form larger themes, (4) reviewing the themes generated, (5) naming and finalizing the themes, and (6) reporting the final themes (Ahmed et al., 2025; Maguire & Delahunt, 2017). This method was used after the researchers transcribed the audio from the interview recordings. By adhering to the six steps of thematic analysis, a common viewpoint among the high school teachers' responses emerged, allowing for the reporting of themes that could be visualized with pieces of evidence available.

For reporting purposes, the teachers recruited were assigned with the pseudonym TOX, which X stands for the number assigned to the high school teachers.

Ethical Considerations

Prior to conducting the study, the researchers had sought ethical approval from the researchers' institution of study. An approval to conduct the research was also sought from Zhejiang Art Vocational College. Throughout the conduct of the study, the participants were informed of their rights and the decision to participate and withdraw from the study were completely voluntary with no further actions taken. At the same time, their anonymity was ensured, with corresponding pseudonyms assigned.

Findings

Based on the thematic analysis conducted, the researchers were able to identify two major themes with their corresponding sub-themes. Through the guidance provided by the Multimodal Learning Theory, this section informs how affordances and constraints of pedagogy are experienced by these high school teachers, especially in terms of integrating the visual and textual components in art classrooms. Figure 1 displays the figure illustrating

the major themes and their corresponding sub-themes. According to the figure, the first major theme obtained was “Perceptions,” which was supplemented by the sub-themes of “Pedagogical Challenges” and “Specialized Training”; the second major theme emerged after the teachers shared their perceptions of using AI-assisted visual text multimodal, which is “Recommendations” that was supported by the sub-theme of “Practical Suggestions.”

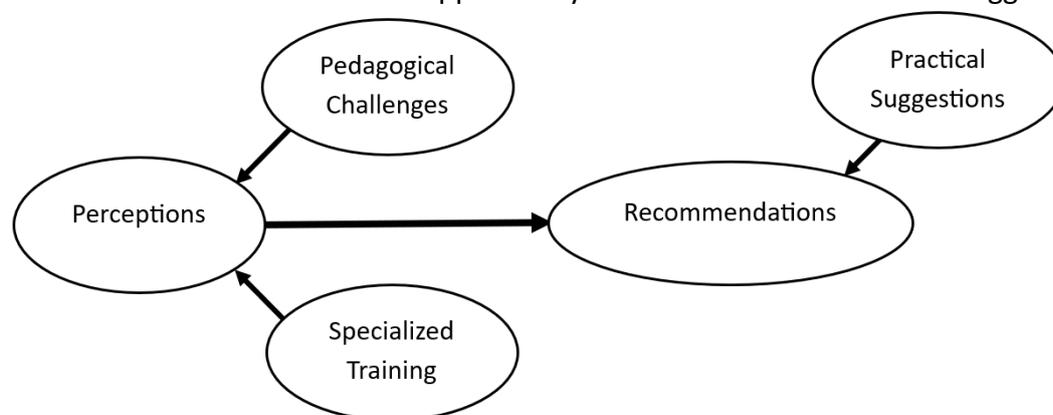


Figure 1 Themes Identified and Obtained

Theme 1: Perceptions

The first major theme obtained from the five high school teachers was the perceptions towards using AI-assisted visual text multimodal tools in arts classrooms. In the perceptions shared by these high school teachers, while they perceived artificial intelligence positively as a useful tool, they also shared some concerns of using AI when it came to pedagogical matters and their readiness.

Sub-Theme: Pedagogical Challenges

The major concern of using artificial intelligence among the teachers was aligning AI with the curriculum standards and learning goals in classrooms. These teachers shared the consensus that AI was useful in generating materials that were visually appealing, attractive, and interesting. However, these teachers also noted that the impressive artistic concepts did not frequently conform to the learning goals outlined in the syllabus, neither did they fully conform to the intended curriculum standards. Consequently, while these teachers agreed the usefulness of using AI in arts lessons, time was wasted in discarding less useful contents generated by AI instead. The following exemplifies the quotes extracted from the interview.

T01: The biggest difficulty lies in determining what the technological tools play in the classroom. No matter which way I choose to determine, I need to redesign the course structure. I'm not young anymore, and this is very time-consuming.

T04: It's like a separate classroom activity rather than a thread. I feel a sense of disconnection and a waste of time.

At the same time, the teachers also expressed their concerns of the possible cognitive overload faced by the students, which was a result of the poorly generated visual content from AI. Although the high school teachers believed the visual contents generated were interesting and supportive of the students' learning process, the teachers were concerned that the visual contents might not meet the intended quality as AI was still developing and improving. At the same time, these teachers were concerned that the students might end up

getting distracted from real learning, having glued to the interesting features of AI to generate various contents unrelated to lessons.

Furthermore, some teachers were also concerned that the features of using AI might lead students to display behaviors that were unrelated to classroom learning. Although they believed independent learning should be encouraged, establishing a proper boundary as to exactly when and how to use AI only for lesson-related purposes was not easy. While AI was still considered effective, students would still be distracted if the teachers could not control the students' classroom behaviors.

Sub-Theme: Specialized Training

A part of the sub-theme that contributed to the main theme of "Perceptions" is "Specialized Training." Specifically, the teachers believed that specialized training on the usage of AI-assisted tools was necessary due to the perceived belief of inadequate preparation they had had. These teachers also emphasized that if training could be provided, they would be able to make use of AI more meaningfully, especially in terms of inserting prompts, designing the contents, evaluating and ensuring the quality of generated outputs, and integrating the outputs into the lessons that could meet the objectives. The following exemplifies the quotes extracted from the interview.

T01: The training for our teachers cannot be merely about learning operations. We are more interested in understanding the usage related to the knowledge of the teaching content and how to use this to achieve our teaching goals.

T02: We only learned how to use it, but not how to teach it. If I could see very standard lesson plans through specific training, I might have gained a lot.

T05: Although AI can give me relatively comprehensive student reports, I have some difficulty understanding them. How can these data be used? I would hesitate unless some specific guidance is provided.

Nevertheless, the current high school teachers, as reported in the interviews, believed that specialized training was necessary to minimize the time spent on trial and error on the usage of AI. Once training could be provided, the teachers believed they would be able to use AI more effectively in arts lessons.

Theme 2: Recommendations

The second theme was built on the perceptions and experiences of using AI-assisted visual text multimodal tools shared by these teachers. During the interviews, these teachers expanded on their responses by providing recommendations of how to improve AI usage in arts classroom and the education reality they were in with their students.

Sub-Theme: Practical Suggestions

Contributing to the main theme of "Recommendations" was the "Practical Suggestions" given by the five high school teachers interviewed. One of the key recommendations shared by the teachers was to provide a structured professional development program, focusing on training teachers to use AI as a form of pedagogy instead of merely relying AI as an operational tool. Given the significance of using prompts to elicit AI to produce the necessary content for educational purposes, the teachers suggested a training program that could guide them to

deliver effective prompts in generative AI. At the same time, these teachers suggested that the training should help them to elicit prompts that allowed contents generated to match with the curriculum standards and learning goals in arts lessons.

T01: Regular training sessions, not just one-off workshops, would be ideal to help us deliver effective prompts.

At the same time, these teachers recommended that the use of AI should not deter teachers away from being the central role in classrooms. Teachers should still be the instructors who provide guidance to the students, and that AI was stressed to only be the instructional tool than being the agent of teaching itself. Simply, the responses by these teachers indicated that the human guidance was still necessary, and AI was used as a means to facilitate learning rather than making AI the full focus of being the instructors in classrooms.

T05: We should reduce the reliance on these apps. We are still the main ones providing guidance to our students.

Furthermore, these teachers also mentioned recommending institutional support as the key to achieving effective implementation of AI in arts lessons. Specifically, the teachers believed that if schools could provide and invest in the technological infrastructure, especially in AI, alongside providing and supplementing them with the necessary guidance to use AI in classrooms, the teaching and learning processes would be more effective. At the same time, the teachers also mentioned if technological support was made available, the teachers could quickly seek technical assistance without the need to spend ample time troubleshooting on technologies that might be beyond their comprehension.

Discussion

Undoubtedly, findings from the interviews with the five high school teachers in the selected college have provided insightful details of the state of using AI in teaching arts. Generally, results from the interviews imply the recognition of the potential benefits of using the AI-assisted visual text multimodal in arts classes, especially in terms of providing visual learning to engage students. The results also strongly imply that AI would be very useful in effective pedagogical implementation, subsequently enhancing potential professional aspects and achieving institutional goals.

The belief that visual contents and textual contents could be aligned to deliver classroom instructions confirms the dual coding processes in the Multimodal Learning Theory. While it is proven with evidence that visual and textual contents can assist students in retaining and comprehending incoming knowledge (Christoforou, 2025; Tan et al., 2025), the concerns of cognitive overload due to the bombardment of visual information from generative AI are a valid form of limitations of using AI in classes. This is consistent with Sweller's (1988) on cognitive load, wherein over bombarding the students with visuals may be overwhelming. The teachers' response thus reinforces the need for a critical evaluation of the current use of AI in educational settings, alongside addressing how this advanced technology can be effectively adapted.

Nevertheless, despite the concerns and challenges shared by these five high school teachers, their calls for specialized training are a positive sign of readiness to embrace AI for classroom teaching and learning. The current action that should be taken is to provide the necessary

training for the teachers, thereby reducing the obstacles that these teachers face when they are using AI. At the same time, providing relevant training programs, especially when this is hosted by the institution or ministry, would greatly assist the teachers to effectively integrate AI into classroom lesson planning. One of the other benefits of providing such training is that the teachers can also develop the necessary technological competency, especially in terms of using AI (Cohn et al., 2024; Tang & Leong, 2025).

Another positive takeaway from the high school teachers' responses in the interviews is the continuation of emphasizing teachers as the main and central figures of teaching in classrooms. Instead of seeing AI as a new instructor, the high school teachers' responses strongly indicate a teacher-and-AI collaboration over the reliance on AI to conduct lessons. Further interpretation of the teachers' responses also shows that teachers are irreplaceable in the teaching and learning context, and AI can serve as a supporting tool to enhance learning for students while assisting the teachers. The teachers' voices in this study, then, effectively highlight the experiences these teachers face when it comes to using AI in classrooms.

Thus, for successful integration of the AI-assisted visual text modal in classrooms to occur, there must be efforts to balance the usage and training to ensure instructions are controlled without experiencing the frustration of using AI effectively. When training is provided through institutional support, alongside making technical assistance available, the gap of the theoretical application of using AI and its actual usage in classrooms can be bridged.

Conclusion

Exploring the high school teachers' experiences of using AI-assisted visual text multimodal through their perceptions and recommendations is the main aim of this study, specifically within the Chinese education context. Through the use of a qualitative exploratory research design, the five high school teachers' voices are effectively presented, allowing a full presentation of valuable insights on the state of using AI in the selected educational setting.

At its core, this study provides a valuable insight on the actual experiences faced by these teachers in contemporary AI usage in classrooms, transcending theoretical discussions and potential practical uses. These insights demonstrate the effectiveness of combining visual and textual elements in enhancing students' art classroom learning experiences, thereby increasing student engagement and motivation. The use of AI also assists teacher to develop materials and contents that are effective in facilitating the learning process for the students. However, it is also evident in the teachers' 'responses that careful attention is necessary, specifically in terms of ensuring a structured use of AI that does not deviate both teachers and students from achieving meaningful learning and realizing learning goals. It is also crucial to note that the used of AI must assist teachers to align the content generated by AI with the curriculum standards outlined by the Chinese education ministry. Specialized training is thus necessary, as AI-assisted visual text multimodal and AI in general can be effectively adapted and implemented in classrooms is best achieved through such an intervention.

Theoretically, the Multimodal Learning Theory is effectively supported by the findings from the interview responses in this study. Learning is best achieved in classroom when visual and textual components are integrated together rather than treating them as individual learning components (Christoforou, 2025; Tan et al., 2025). Still, even though this method of learning

is effective, it is also evident that cognitive overload through the bombardment of visuals, regardless of high quality or low quality, might affect the students' learning. Another potential downside is the distraction caused by the advanced features offered by AI, thereby needing practical interventions to ensure these downsides are addressed.

In addition to the theoretical contributions, it is also practical to address that continuous and ongoing professional development, as a means of support from the institution, is necessary to assist teachers in enhancing their pedagogical capabilities in classes. When guidelines and assistance are made available, teachers can effectively implement AI in classrooms, allowing quality learning to occur.

Still, this study is not without limitations that require addressing. Since the present study only involved a small group of teachers, generalization of findings and transferability of findings are limited. Expanding the number of participants and including other schools in similar research would help enrich the findings to provide a richer description on the state of using AI. At the same time, future studies related to this field may consider researching on how AI-assisted visual text multimodal can help achieving specific learning outcomes and the curriculum standards.

To conclude, AI-assisted visual text multimodal is a promising tool to boost student learning, and the teachers' experiences, perceptions, and recommendations should be accounted to enhance the quality of teaching and learning in the Chinese educational context. This study has thus informed a balanced view of understanding AI, its contribution to the arts classes, and potential suggestions for future research and intervention.

Acknowledgement

The researchers declare no conflicting interests in conducting the study. The entire study is conducted solely by the researchers themselves.

References

- Ahmed, S. K., Mohammed, R. A., Nashwan, A. J., Ibrahim, R. H., Abdalla, A. Q., Ameen, B. M. M., & Khahir, R. M. (2025). Using thematic analysis in qualitative research. *Journal of Medicine, Surgery, and Public Health*, 6. <https://doi.org/10.1016/j.glmedi.2025.100198>
- Al-Jarf, R. (2024). *Multimodal teaching and learning in the EFL college classroom*. Paper presented at the 2nd International Positive Schools and Well-Being Congress. <https://files.eric.ed.gov/fulltext/ED662915.pdf>
- Alsohaimi, M., Albahiri, M., & Alhaj, A. (2025). Addressing and managing artificial intelligence (AI) challenges and opportunities in elementary education in Saudi Arabia: An in-depth consideration. *Educational Process: International Journal*, 17. <https://doi.org/10.22521/edupij.2025.17.324>
- Arifin, K., Sirih, M., Munir, A., Jahidin, Sabilu, M. (2025). The influence of multimodal learning strategies on perspective biology teachers' literacy-numeracy learning outcomes. *Eurasia Journal of Mathematics, Science and Technology Education*, 21(1). <https://doi.org/10.29333/ejmste/15802>
- Chen, Q., Liu, S., Huang, K., Wang, X., Ma, X., Zhu, J., & Peng, Z. (2024). *Ret assist: Facilitating vocabulary learners with generative images in story retelling practices*. Paper presented

- at the Proceedings of the 2024 Designing Interactive Systems Conference, 1432-1446.
<https://dl.acm.org/doi/10.1145/3643834.3661581>
- Christoforou, M. (2025). Gen AI-assisted multimodal meaning design: exercising a pedagogic metalanguage of transposition. *Pedagogies: An International Journal*.
<https://doi.org/10.1080/1554480X.2025.2522884>
- Cohn, C., Snyder, C., Fonteles, J. H., Ashwin, T. S., Montenegro, J., & Biswas, G. (2024). A multimodal approach to support teacher, researcher and AI collaboration in STEM+C learning environments. *British Journal of Educational Technology*, 56(2), 595-620.
<https://doi.org/10.1111/bjet.13518>
- Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson Education, Inc.
<https://www.pearson.com/us/higher-education/program/Creswell-Educational-Research-Planning-Conducting-and-Evaluating-Quantitative-and-Qualitative-Research-plus-My-Lab-Education-with-Enhanced-Pearson-e-Text-Access-Card-Package-6th-Edition/PGM335066.html>
- Edmonds, W. A., & Kennedy, T. D. (2017). *An applied guide to research designs: Quantitative, qualitative, and mixed methods* (2nd ed.). SAGE.
<https://methods.sagepub.com/book/an-applied-guide-to-research-designs-2e>
- Guan, L., Zhang, E. Y., & Gu, M. M. (2025). Examining generative AI-mediated informal digital learning of English: A mixed-methods study. *ReCALL*, 37(1), 1-22.
<https://www.cambridge.org/core/journals/recall/article/examining-generative-aimediated-informal-digital-learning-of-english-practices-with-social-cognitive-theory-a-mixedmethods-study/F9B7ACD7BFEC6A3864463DD7F228AC27>
- He, S., & Zhang, M. (2025). *Boosting artistic creativity in Chinese primary school children via AI technology*. Paper presented at the Proceedings of the 2025 2nd International Conference on Digital Systems and Design Innovation, 131-137.
<https://doi.org/10.1145/3759275.3759295>
- Hunter, D., McCallum, J., & Howes, D. (2019). Defining exploratory-descriptive qualitative (EDQ) research and considering its application to healthcare. *Journal of Nursing and Health Care*, 4(1). <http://eprints.gla.ac.uk/180272/>
- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
<https://archive.org/details/multimodaldiscou0000kres>
- Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*.
<https://ojs.aishe.org/index.php/aishe-j/article/download/335/553/1557>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Ng, D. T. K., Chan, E. K. C., & Lo, C. K. (2025). Opportunities, challenges and school strategies for integrating generative AI in education. *Computers and Education: Artificial Intelligence*, 8. <https://doi.org/10.1016/j.caeai.2025.100373>
- Olawale, S. R., Chinagozi, O. G., & Joe, O. N. (2023). Exploratory research design in management science: A review of literature on conduct and application. *International Journal of Research and Innovation in Social Science*, 7(4).
<https://doi.org/10.47772/IJRISS.2023.7515>
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)

- Tan, X., Xu, W., & Wang, C. (2025). *Voice in AI-assisted multimodal texts: What do readers pay attention?* *Computers and Composition*, 75. <https://doi.org/10.1016/j.compcom.2025.102918>
- Tang, Y. F., & Leong, W. Y. (2025). Integrating AI-generated art styles into traditional illustration design teaching course. *Eurasia Journal of Mathematics, Science and Technology Education*, 21(7). <https://doi.org/10.29333/ejmste/16602>
- Vatral, C., Biswas, G., & Goldberg, B. (2023). *A theoretical framework for multimodal learner modeling and performance analysis in experiential learning environments*. Paper presented at the International Conference on Artificial Intelligence in Education. https://ceur-ws.org/Vol-3484/AIED-GEL23_paper_7_CEUR.pdf
- Zhang, Y. (2025). The impact and challenges of generative artificial intelligence on the arts education. *Curriculum and Teaching Methodology*, 8(5). https://clausiuspress.com/assets/default/article/2025/07/12/article_1752376388.pdf