STEM Education Review

REVIEW ARTICLE

From the "Metaverse Classroom" to Intelligent Immersive Learning Systems: A Four-Layer Framework for Foreign-Language Education in Chinese Colleges

Zhishan Liu[1,2,*], Nur Aimi Nasuha Burhanuddin[1], Nor Aniza Binti Ahmad[1]

[1]Faculty of Educational Studies, Universiti Putra Malaysia, 43400 Selangor, Malaysia
[2]Faculty of Educational Studies, Hainan College of Foreign Studies, 571321 Wenchang, China

ABSTRACT

The digital transformation of foreign-language education in higher vocational colleges is shifting from whether to adopt virtual environments to how intelligent technologies can sustain support for occupational language learning. Early research on the educational metaverse emphasized three-dimensional representation, immersive interaction, and collaborative resource development. However, virtual space alone does not automatically produce reliable language-learning outcomes. Recent advances in generative artificial intelligence, conversational agents, extended reality, and learning analytics now enable virtual environments to generate tasks in real time, simulate multi-party workplace communication, provide individualized feedback, and retain process-based evidence of learning. Drawing on situated learning, engagement theory, and the Cognitive-Affective Model of Immersive Learning, this conceptual article proposes a four-layer framework comprising (a) a scenario layer organized around occupational tasks, (b) an intelligent interaction layer centered on multi-turn dialogue, (c) a learning-evidence layer based on process data, and (d) a governance layer addressing safety, ethics, and equity. The framework is translated into a course-design model that aligns occupational task chains, learning task chains, and assessment evidence chains. A phased implementation pathway is also proposed, focusing on low-cost pilots, teacher capacity development, college-industry collaboration, and risk governance. The central argument is that technological choices should remain subordinate to occupational competence goals: immersion can generate sustainable learning gains only when it is integrated with authentic tasks, high-quality feedback, and structured reflection.

Keywords: Foreign-language education; Higher vocational education; Generative artificial intelligence; Extended reality (XR); Intelligent agents

INTRODUCTION

Foreign-language-focused higher vocational colleges are expected to teach linguistic knowledge while preparing students for occupational communication across cross-border business, tourism services, international communication, translation, and other internationally oriented fields. Their central pedagogical challenge is therefore not simply a shortage of digital resources. Rather, a persistent gap remains between classroom language tasks and workplace performance: students may complete textbook exercises successfully yet struggle to communicate when information is incomplete, role relationships are complex, and feedback shifts in real time. Liu and Liu (2025) responded to this problem by examining how metaverse technologies might reshape instructional scenarios, inter-

***Corresponding Author:**
Email: liuzhishan333@163.com
Received: 14 July 2025; Revised: 19 October 2025; Accepted: 20 December 2025
https://doi.org/10.67229/stemer.26080003

action, and educational ecosystems in foreign-language vocational education. Their analysis proposed a shift from two-dimensional classrooms to three-dimensional environments and from one-way instruction to immersive participation. That argument remains instructive, but the technological context has since changed. Virtual reality is no longer only a display medium; generative artificial intelligence (GenAI) can now operate within virtual scenarios as a customer, colleague, examiner, or learning partner whose responses change dynamically.

Technological visibility, however, should not be mistaken for educational effectiveness. Systematic reviews have identified persistent problems in educational-metaverse research, including conceptual overextension, limited empirical evidence, and insufficient attention to cost and privacy risks (Camilleri, 2024; Tlili et al., 2022). Studies of language learning generally report that extended-reality technologies can support motivation, interaction, and selected language skills (Chen et al., 2022; Huang et al., 2021). Nevertheless, a recent review has also documented dizziness, fatigue, cognitive overload, and inadequate feedback design as barriers to effective immersive learning (Rahmanu & Molnár, 2026). Foreign-language vocational education should therefore move beyond the technological narrative of "building a metaverse classroom" toward the design of intelligent immersive learning systems that are assessable, governable, and sustainable.

This article develops a conceptual framework based on prior research and policy documents. It reports no new experimental data and does not treat technological affordances as verified instructional effects. Three questions guide the discussion. First, how does incorporating GenAI into virtual environments change the mechanisms of occupational language learning? Second, how can scenarios, interaction, evidence, and governance be organized within an integrated framework? Third, how can higher vocational colleges with varying levels of infrastructure and staff capacity implement the framework in stages?

FROM SPATIAL DISPLAY TO TASK GENERATION: A SHIFT IN THE TECHNOLOGICAL LOGIC

Educational Value Lies in the Continuous Generation of Occupational Tasks

The educational metaverse is commonly described as a learning space characterized by the integration of physical and virtual environments, persistent identities, multi-user collaboration, and interconnected data (Liu et al., 2022; Zhai et al., 2022). From an educational-technology perspective, however, its value should be judged by how effectively it improves learning processes and resources, rather than by the number of technologies it combines (He, 2005). From an artificial-intelligence perspective, the more consequential development is the transformation of a virtual environment into a system that can respond to learner behavior: characters can interpret language, events can change in response to decisions, task difficulty can be adjusted dynamically, and the learning process can be recorded continuously (Hwang & Chien, 2022). The appropriate unit of instructional design is therefore not an isolated three-dimensional scene, but a complete task with an objective, roles, constraints, unexpected events, and performance standards.

Consider an English-for-cross-border-e-commerce course. A visual reconstruction of an exhibition hall or livestream studio provides only a background. An intelligent immersive task, by contrast, can require learners to analyze product information, explain a product to customers from different cultural backgrounds, modify their language when challenged about returns or regulatory compliance, and submit written follow-up communication. Artificial intelligence is valuable in this setting because it can generate differentiated follow-up questions and controlled changes in task conditions, not because it can complete the interaction for the learner. Consequently, the technological system shifts from a content player to a task generator, while the teacher's role shifts from explaining resources to designing tasks, interpreting evidence, and safeguarding educational values.

Immersion Produces Learning Gains Only Through Cognitive and Affective Mechanisms

Situated learning theory holds that knowledge derives meaning from participation in a community of practice and that competence develops through progressively fuller participation in authentic or simulated activity (Lave & Wenger, 1991). Engagement theory similarly emphasizes collaboration, creation, and authentic tasks as foundations for technology-supported learning (Kearsley & Shneiderman, 1998). The Cognitive Affective Model of Immersive Learning further explains that presence and agency can influence interest, self-efficacy, and cognitive load, but that greater immersion does not necessarily yield better learning (Makransky & Petersen, 2021). When a scenario contains excessive information or imposes a heavy operational burden, learners may allocate their limited attention to using the device rather than completing the language task.

Foreign-language courses in higher vocational colleges should therefore adopt a principle of proportionate immersion. Knowledge analysis that can be completed adequately with two-dimensional materials does not require a head-mounted display. Extended reality becomes distinctive only when spatial relations, role configurations, embodied action, or workplace pressure forms part of the competence being developed. Existing reviews also show that many virtual-reality applications in higher education emphasize usability and novelty while offering limited evaluation of learning outcomes, retention, and transfer (Radianti et al., 2020). Technology should be evaluated according to whether it improves consequential learner behavior, not according to the sophistication of the equipment.

THREE STRUCTURAL CHALLENGES IN FOREIGN-LANGUAGE VOCATIONAL EDUCATION

Textbook Scenarios Are Complete, Whereas Workplaces Are Uncertain

Traditional materials typically provide clear roles, standard expressions, and a single correct answer. Real-world workplaces, by contrast, involve missing information, conflicting interests, cultural differences, and time pressure. A learner who performs fluently with a fixed script may be unable to proceed when an interlocutor changes the question or refuses to cooperate. The purpose of a virtual scenario is therefore not merely to increase visual realism. It is to introduce controlled uncertainty so that learners can practice clarification, negotiation, communication repair, and decision-making in a comparatively low-risk environment.

More Interaction Does Not Necessarily Mean Better Feedback

Conversational agents can increase opportunities for oral practice and may reduce the anxiety some learners experience when speaking with another person. A systematic review found promising applications of AI chatbots for speaking practice, pronunciation support, engagement, and confidence, but also concluded that the evidence base remains at an early stage, and that long-term outcomes and differences among learners require further investigation (Du & Daniel, 2024). Research on ChatGPT's affordances for language education similarly emphasizes that teachers and learners need appropriate digital competencies if the technology is to be used ethically and effectively (Kohnke et al., 2023). Broader guidance also warns of misinformation, bias, overreliance, and threats to academic integrity (Miao & Holmes, 2023).

High-quality feedback should address at least four dimensions: task completion, informational accuracy, linguistic appropriateness, and strategic effectiveness. Artificial intelligence may provide initial annotations and alternative expressions, but the course team must define the central assessment criteria, and teachers or industry mentors must verify consequential facts and professional standards. Learners should also be required to explain, revise, and perform again after receiving feedback. Without this cycle, frequent interaction is unlikely to produce durable improvement.

Process Data Are Abundant, but Assessment Remains Dominated by Final Scores

Intelligent immersive environments can record pauses, reformulations, requests for assistance, role changes, task pathways, and the number of revisions. These data create opportunities for formative assessment, but they can also encourage the misconception that more data necessarily produce more scientific assessment. Meaningful learning evidence must be explicitly mapped to competence standards. Institutions must ask what a data point represents, whether a teacher can interpret it, whether it disadvantages particular groups, and whether learners can inspect and challenge the resulting judgment.

China's 2025 policy on educational digitalization calls for the use of big data and artificial intelligence to support process, value-added, and comprehensive assessment while simultaneously emphasizing trustworthy, controllable, and safe development and the regulation of AI applications in educational institutions (Ministry of Education of the People's Republic of China et al., 2025). Innovation in assessment and governance of data should therefore proceed together.

A FOUR-LAYER FRAMEWORK FOR INTELLIGENT IMMERSIVE FOREIGN-LANGUAGE LEARNING

Scenario Layer: Organizing the Environment Around Occupational Events

The scenario layer addresses what work learners must complete and under which conditions. Course teams should first identify frequent, high-risk, or communication-intensive events through program specifications and occupational research. They can then determine whether a task requires text, video, a desktop three-dimensional environment, or immersive equipment. A complete scenario should specify task objectives, role relationships, information resources, constraints, unexpected events, and expected deliverables. Authenticity does not require a literal reproduction of physical space; it requires the preservation of variables that materially affect workplace judgment.

Intelligent Interaction Layer: Giving Roles Bounded Generative Capacity

The intelligent interaction layer specifies how the system responds to the learner. Each role should have a stable identity, a defined knowledge boundary, a communicative purpose, and limits on emotional expression. In a hotel-complaint scenario, for example, a guest agent may express dissatisfaction and ask follow-up questions but should not invent company policy. In a business-negotiation scenario, a customer agent may modify conditions, but any change should remain within parameters established by the teacher. System prompts, knowledge bases, and safety rules should be treated as version-controlled course resources. Teachers should also be able to intervene manually so that erroneous feedback is not amplified across multiple conversational turns.

Learning-Evidence Layer: Converting Behavioral Traces Into Interpretable Assessment

The learning-evidence layer addresses how genuine progress can be recognized. Assessment may combine four categories of indicators. The first is language performance, including accuracy, fluency, register, and intercultural appropriateness. The second is task performance, including information gathering, problem-solving, and the quality of the final deliverable. The third is process performance, including the uptake of feedback, strategic adjustment, and contributions to collaboration. The fourth is transfer performance: the ability to complete a comparable task with a new role, product, or constraint. Meta-analytic evidence suggests that virtual-reality-assisted language learning can produce moderate linguistic and affective benefits (Chen et al., 2022). College-level assessment should nevertheless examine retention and transfer rather than relying only on immediate post-lesson scores.

Governance Layer: Establishing Safety, Equity, and Human Agency as System Constraints

Governance is not an add-on implemented after a project has been completed; it defines the operating boundaries of the other three layers. UNESCO guidance recommends a human-centered approach to GenAI in education, the protection of data privacy, and the pedagogical and ethical validation of tools (Miao & Holmes, 2023). UNESCO's AI Competency Framework for Teachers identifies five interrelated dimensions: a human-centered mindset, the ethics of AI, AI foundations and applications, AI pedagogy, and AI for professional learning (Miao & Cukurova, 2024). Higher vocational colleges should accordingly establish procedures for tool approval, data minimization, content verification, learner notification, human appeal, accessible alternatives, and withdrawal. No learner should be excluded from core learning because they cannot use expensive equipment, experience physical discomfort, or decline to submit biometric data.

ALIGNING OCCUPATIONAL TASKS, LEARNING TASKS, AND ASSESSMENT EVIDENCE

Deriving the Learning Task Chain from the Occupational Task Chain

Course development should begin by mapping the complete sequence of a representative workplace event and then decomposing it into teachable units. An international exhibition-reception task, for example, may include clarifying needs, presenting information, giving directions, managing an unexpected problem, and conducting follow-up communication. Each unit should define its inputs, observable behaviors, quality standards, and common errors. Teachers can then decide which parts should be performed with an intelligent agent, which require human collaboration, and which knowledge must be taught explicitly before learners enter the scenario.

Building an Evidence Chain Through Preparation, Performance, Debriefing, and Reperformance

A complete learning activity can be organized into four stages. During preparation, learners analyze the task, verify terminology and professional rules, and explain the communication strategy they intend to use. During performance, they enter a virtual or blended scenario and complete the task with intelligent agents and peers. During debriefing, they review video, dialogue records, and assessment rubrics to identify consequential decisions and language problems. During reperformance, roles or constraints are changed so that the transfer of feedback can be examined. GenAI may help generate new scenarios of comparable difficulty, but teachers must approve the final tasks and scoring rules.

Using Assessment Results to Improve Instruction Rather Than Produce Finer Rankings

The primary purpose of learning analytics is to identify which task, which form of support, and which learners require adjustment. At the class level, teachers can identify recurring points of failure. At the course level, teams can compare cognitive load and transfer across scenarios. At the individual level, learners should receive a small number of actionable recommendations. For automated assessment, institutions should conduct regular human audits and test whether a model introduces systematic bias related to accent, expressive style, or cultural background. Course specifications should also state how long assessment evidence will be retained, who may access it, and how it can be deleted.

A PHASED IMPLEMENTATION PATHWAY

Beginning With a Low-Cost, Task-Intensive Minimum Viable Design

Institutions do not need to begin by constructing a large-scale "metaverse campus." A more sustainable approach is to select one high-value task within a course and use ordinary computers or mobile devices to implement a desktop scenario and intelligent dialogue. The first pilot should test the task design, feedback cycle, and assessment rubric. Head-mounted equipment should be added only when spatial interaction materially affects the learning outcome. Cost-benefit research likewise indicates that the appeal of immersion should be weighed against the costs of equipment, maintenance, training, safety, and continuing content development (Camilleri, 2024).

Redirecting Teacher Development From Tool Operation to Pedagogical Judgment

Teacher development should not be measured by the number of platforms an instructor can operate. It should cultivate five capabilities: identifying appropriate tasks, designing roles and constraints, verifying generated content, interpreting process data, and addressing ethical and safety concerns. Collaborative lesson design, peer review, and small-scale classroom validation can help teachers develop transferable design knowledge in authentic courses. Teachers must also retain the ability to reject a technology. A simpler instructional method should be selected when a tool does not improve consequential learning behavior or when its data practices are opaque.

Replacing Industry "Content Provision" With Joint Definition of Tasks and Standards

College-industry collaboration should shift from the donation of resources to the joint definition of occupational events, performance standards, and exceptional conditions. Industry mentors can provide de-identified cases, review the language and professional rules embedded in an intelligent agent, and evaluate samples of learner work. Colleges remain responsible for converting workplace experience into tasks that are learnable, repeatable, and assessable. UNESCO-UNEVOC's guidance on AI in technical and vocational education similarly frames AI integration as an institution-wide process encompassing governance, curricula, teaching, assessment, capacity development, and monitoring, rather than as the isolated adoption of a tool (Yang et al., 2026).

Developing a Risk Register Alongside Every Instructional Pilot

Every pilot should complete a risk register before teaching begins. At minimum, it should address generated errors, intellectual property, personal information, model bias, overreliance, physical discomfort, equipment accessibility, and vendor lock-in. High-risk activities require human review, while unacceptable risks require explicit stop conditions. Institutions should also explain to learners which content has been generated by AI, which data are recorded, how much weight automated assessment carries, and how a decision can be challenged. Specific governance measures allow teachers and learners to explore intelligently within clearly defined boundaries.

CONCLUSIONS

The next stage of digital transformation in foreign-language vocational education should not be organized around whether an institution has "built a metaverse." It should return to the central purpose of vocational education: enabling learners to use appropriate language to complete authentic tasks in complex settings and to improve continuously through evidence. Gen-AI expands the variability and feedback capacity of virtual scenarios, while extended reality enhances spatial presence. Neither technology is likely to deliver stable educational value unless the scenario, interaction, evidence, and governance layers work together.

The framework recommends starting with bounded course pilots instead of large infrastructure projects. Each pilot should focus on one occupational event, defining the language and outcomes to observe, and clarifying which decisions remain with teachers. Evidence should be gathered during preparation, performance, debriefing, and reperformance to assess if immersive features enhance learning beyond simpler media. Evaluation should include language scores, retention, transfer, learner workload, access, teacher time, and technical costs.

This evidence would allow colleges to distinguish pedagogically necessary immersion from technological novelty and to revise weak components before expansion. Industry partners can contribute task standards and exceptional cases, while teachers retain responsibility for instructional judgment and student support. This staged approach also makes governance testable: privacy protections, appeal procedures, accessible alternatives, and human review can be examined during the pilot rather than added after implementation. In this way, scaling becomes an evidence-informed decision about educational value, not a symbolic measure of digital modernization. To support comparison across settings, future pilots should report the institutional context, learner characteristics, task complexity, technology configuration, teacher preparation, implementation costs, and governance arrangements, thereby allowing researchers to interpret outcomes responsibly and identify which combinations of pedagogical and technological conditions are transferable across vocational language programs in diverse contexts.

This article proposes a four-layer conceptual framework and translates it into an aligned course-design and implementation pathway. Its contribution is conceptual rather than empirical. Thus, institutions should begin with small, auditable pilots that connect workplace tasks, learner evidence, teacher oversight, and iterative course improvement. Future research should therefore adopt longitudinal, cross-course, and college-industry designs; compare the cost-effectiveness of different levels of immersion and different feedback mechanisms; and examine skill transfer, learning equity, and teacher professional judgment. Technologies can change rapidly, but occupational competence standards and human agency should remain the anchors of system design.

DECLARATION

Source of funding

The authors received no specific funding for this research, its analysis, or the preparation of this manuscript.

Conflict of interest

The authors declare no competing interest.

Data availability statement

No new empirical data were generated or analyzed for this conceptual article.

REFERENCES

Camilleri, M. A. (2024). Metaverse applications in education: A systematic review and a cost-benefit analysis. *Interactive Technology and Smart Education, 21*(2), 245–269. https://doi.org/10.1108/ITSE-01-2023-0017

Chen, B., Wang, Y., & Wang, L. (2022). The effects of virtual reality-assisted language learning: A meta-analysis. *Sustainability, 14*(6), Article 3147. https://doi.org/10.3390/su14063147

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence, 6*, Article 100230. https://doi.org/10.1016/j.caeai.2024.100230

He, K. (2005). On the logical starting point of educational technology]. *E-Education Research*, (11), 3–19.

Huang, X., Zou, D., Cheng, G., & Xie, H. (2021). A systematic review of AR and VR enhanced language learning. *Sustainability, 13*(9), Article 4639. https://doi.org/10.3390/su13094639

Hwang, G.-J., & Chien, S.-Y. (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence, 3*, Article 100082. https://doi.org/10.1016/j.caeai.2022.100082

Kearsley, G., & Shneiderman, B. (1998). Engagement theory: A framework for technology-based teaching and learning. *Educational Technology, 38*(5), 20–23.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Liu, G., Gao, N., Hu, H., & Qin, Y. (2022). Educational metaverse: Characteristics, mechanisms, and application scenarios. *Open Education Research, 28*(1), 24–33.

Liu, Z., & Liu, L. (2025). Exploring educational-technology pathways for foreign-language higher vocational colleges in metaverse scenarios]. *Contemporary Foreign Language Education* (Vol. 12), 68–73.

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*, 937–958. https://doi.org/10.1007/s10648-020-09586-2

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://doi.org/10.54675/ZJTE2084

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Ministry of Education of the People's Republic of China, Cyberspace Administration of China, National Development and Reform Commission, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security, Ministry of Finance, Ministry of Human Resources and Social Security, & National Data Administration. (2025, April 11). *Opinions on accelerating educational digitalization*. https://www.moe.gov.cn/srcsite/A01/s7048/202504/t20250416_1187476.html

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, Article 103778. https://doi.org/10.1016/j.compedu.2019.103778

Rahmanu, I. W. E. D., & Molnár, G. (2026). Virtual reality in English language education: A systematic review of opportunities and challenges in higher education. *Acta Psychologica, 264*, Article 106358. https://doi.org/10.1016/j.actpsy.2026.106358

Tlili, A., Huang, R., Shehata, B., Liu, D., Zhao, J., Metwally, A. H. S., Wang, H., Denden, M., Bozkurt, A., Lee, L.-H., Beyoglu, D., Altinay, F., Sharma, R. C., Altinay, Z., Li, Z., Liu, J., Ahmad, F., Hu, Y., Salha, S., . . . Burgos, D. (2022). Is metaverse in education a blessing or a curse: A combined content and bibliometric analysis. *Smart Learning Environments, 9*, Article 24. https://doi.org/10.1186/s40561-022-00205-x

Yang, C., Wu, W., & Tegelbeckers, H. (2026). *Integrating AI in TVET: A practical guide for institutions*. UNESCO-UNEVOC. https://www.unesco.org/en/articles/integrating-ai-tvet-practical-guide-institutions

Zhai, X., Chu, X., Wang, M., Zhang, Z., & Dong, Y. (2022). Educational metaverse: Innovations and challenges of a new-generation internet-based form of education . *Open Education Research, 28*(1), 34–42.