

ImmersiveCFL: Multi-Agent Learning Path Generation for Scenario-Oriented Chinese Language Learning

Shanghai Yang Zhixuan Wu Yanan Zeng Feixiang He
Chunyang Wang* Weining Qian

East China Normal University, China

{shyang0626, 10225501419, yann_zeng, 51275903103}@stu.ecnu.edu.cn

{cywang, wnqian}@dase.ecnu.edu.cn

Abstract

As online language education increasingly shifts toward personalized and immersive learning, scenario-based learning (SBL) offers a promising way to connect language knowledge with authentic communicative contexts. However, creating or generating immersive scenario-oriented learning materials remains challenging, since systems must interpret complex real-world situations, organize them into pedagogically coherent learning paths, adapt linguistic difficulty to learners' proficiency levels, and support interactive learning beyond static content delivery. In this work, we develop **ImmersiveCFL**, a multi-agent framework that generates scenario-oriented learning paths for enhancing immersive experiences during learning Chinese as Foreign Language (CFL). Given learner-customized scenarios, ImmersiveCFL schedules a pedagogically coherent learning path with knowledge retrieved from a Chinese Education Knowledge Graph, generates multimodal scenario dialogues, and creates task-based interactive simulations with multi-skill exercises. Experimental results show that ImmersiveCFL improves difficulty control, curriculum coherence, and personalization over general LLM-based baselines. An online prototype¹ further demonstrates its potential for goal-driven and interactive CFL learning.

1 Introduction

Recently, Chinese language education has witnessed rapidly growing global demand, driven by expanding economic, cultural, and academic exchanges. According to the 2025 language report² released by Duolingo, Chinese has become the fastest-growing language studied in more than ten countries and the second fastest-growing language studied in the United States. Online learning platforms for Chinese as a Foreign Language (CFL)

*Corresponding author.

¹<http://smartlearn.shuishan.net.cn:8089>

²Duolingo 2025 language report

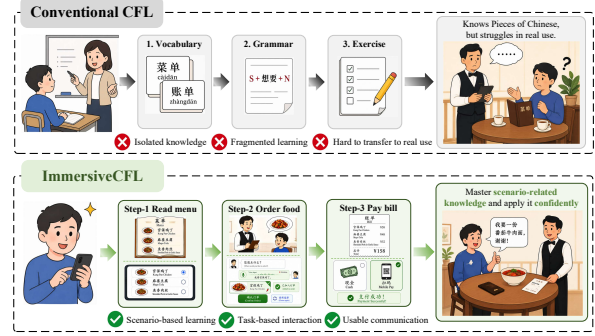


Figure 1: From conventional CFL learning to ImmersiveCFL. Learners acquire scenario-related knowledge and practice it through goal-driven interaction.

have provided scalable access to video lectures, flashcards, listening materials, and exercises.

However, most existing platforms still rely on pre-designed materials and fixed topic sequences, making it difficult to support immersive learning experiences for learners with diverse real-life communicative needs. As shown in Fig. 1, conventional CFL learning typically presents vocabulary, grammar, and exercises as separate instructional units, which hinders learners from transferring acquired knowledge to authentic communication. Situated learning theory (Lave and Wenger, 1991) suggests that language acquisition is deeply embedded in authentic contexts and communicative practices. Scenario-based learning (SBL) (Smith et al., 2018; Misfeldt, 2015) offers a natural basis for immersive CFL learning, which seamlessly injects language knowledge into communicative situations from daily life or professional communication.

However, generating immersive scenario-oriented CFL materials is both labor-intensive and technically challenging in three aspects. First, real-world scenarios are inherently complex and often involve temporal progression, causal dependency, role interaction, and cultural conventions. Without effective scenario decomposition, generated

materials may become fragmented or fail to guide learners from basic expressions to more complex communicative tasks. Second, language learning resources must remain pedagogically appropriate with learners’ language proficiency, such as their Hanyu Shuiping Kaoshi (HSK) (Chinese Testing International, 2026) level. However, general-purpose Large Language Models (LLMs) often lack explicit mechanisms for learner-aware difficulty control and pedagogical sequencing. Third, scenario-oriented CFL systems should support interactive problem-solving simulations in which learners can practice language within dynamic environments by completing concrete goals, making choices, and receiving feedback. To address these challenges, we propose **ImmersiveCFL**, a multi-agent framework that transforms learner-customized scenarios into immersive CFL learning experiences. Given a scenario and learner profile, a **Learning Path Scheduler** decomposes into pedagogically coherent sub-scenarios and retrieves difficulty-appropriate knowledge from a Chinese Education Knowledge Graph. A **Scenario Dialogue Generator** then produces multimodal role-playing dialogues as contextualized learning materials. Finally, an **Interactive Practice Generator** creates task-based HTML interactions and multi-skill exercises, supporting interactive problem-solving simulations. Evaluation experiments show that ImmersiveCFL improves difficulty control, curriculum coherence, and personalization, enhancing immersion during learning.

2 Related Work

2.1 Task-based Language Teaching & Scenario-based Learning

Task-Based Language Teaching (TBLT) and Scenario-Based Learning (SBL) provide the theoretical foundation for contextualized language acquisition. Regarding TBLT, Ellis (Oxford, 2006) established its roots in cognitive-interactionist theories, arguing that acquisition is driven by learners’ engagement in authentic tasks. Long (Long, 2014) further proposed a TBLT framework that prioritizes real-world objectives over discrete grammatical practice, while Willis (Willis, 2021) introduced a structured task cycle in which communicative tasks are followed by form-focused analysis to develop both fluency and accuracy. Concurrently, SBL literature has advanced the understanding of highly contextualized learning. Lave and Wenger (Lave

and Wenger, 1991) proposed situated learning theory to emphasize that learning is embedded in social practice, and Errington (Errington, 2003) introduced formal SBL models that place learners in immersive and narrative-driven environments. These works suggest effective language learning should connect linguistic knowledge with concrete communicative goals and real-world scenarios.

2.2 Adaptive Instructional Planning

Instructional planning has long been regarded as a core component of intelligent tutoring systems. Early studies mainly relied on symbolic representations and rule-based heuristics to align instructional content with learning outcomes (Gagne et al., 1981), organize learning paths through concept maps and state transitions (Peachey and McCalla, 1986), and support adaptive plan repair according to student progress (Vassileva, 1992, 1995).

Recent advances in deep learning and LLMs have introduced new perspectives. Learning Path Recommendation (LPR) aims to generate individualized learning resource sequences (Nabizadeh et al., 2020), including sequence-based methods that construct complete learning paths (Zhou et al., 2018; Chen et al., 2023) and step-based methods that recommend learning elements according to learners’ evolving feedback (Li et al., 2023; Liu et al., 2019). More recently, Multi-Agent Educational Systems (MAES) have leveraged LLM agents for adaptive instructional planning (Lv et al., 2025; Wang et al., 2025; Zhang et al., 2025). However, existing methods mainly focus on resource sequencing or general instructional design, while limited attention has been paid to language education. ImmersiveCFL bridges this gap by employing multiple specialized agents to coordinate scenario understanding, knowledge-grounded planning and immersive practice for personalized CFL learning.

3 Framework Design

In this section, we first formally define the **Scenario-oriented Learning Path Generation** task, and then elaborate on the three-stage generation pipeline of ImmersiveCFL shown in Fig. 2.

3.1 Problem Definition

Formally, we define the *Scenario-oriented Learning Path Generation* task in this work as follows:

$$\mathcal{L} = LPG(S, \mathcal{U}) = \{\ell_i \mid \ell_i = \langle V_i, D_i, E_i \rangle\}_{i=1}^N. \quad (1)$$

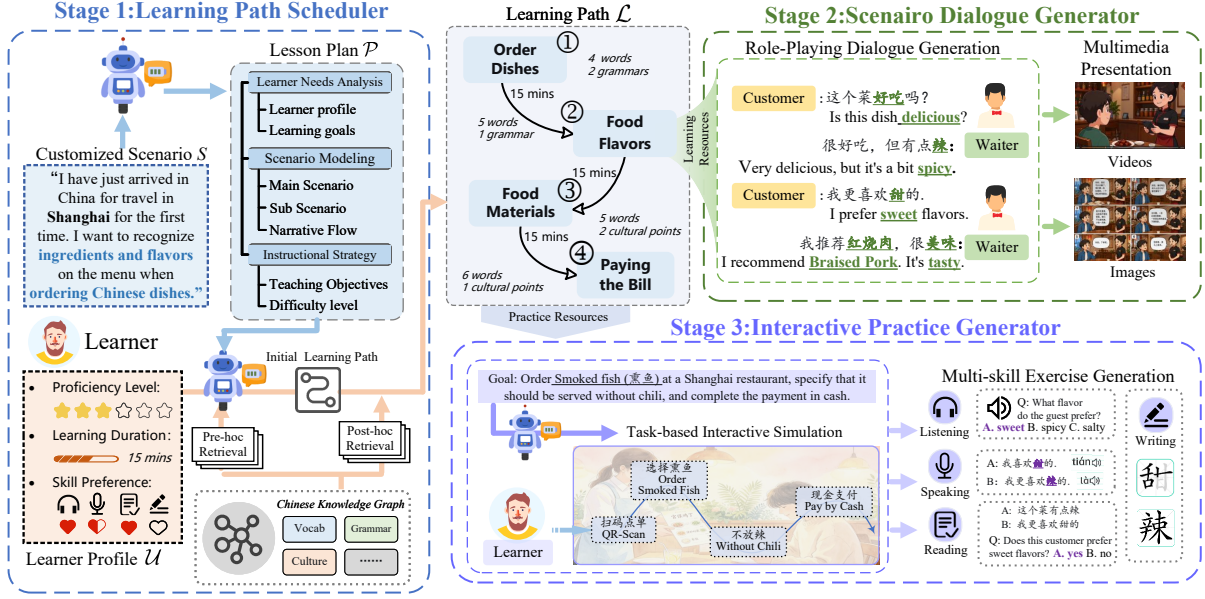


Figure 2: The overall multi-agent generation pipeline of ImmersiveCFL.

Input: S is the learner-provided scenario description, e.g., *Ordering food in Beijing*; $\mathcal{U} = \langle \theta, T, \mathbf{w}_{skill} \rangle$ denotes the learner profile, including proficiency level θ , expected lesson duration T , and skill preference $\mathbf{w}_{skill} \in \mathbb{R}^5$ over *Listening, Speaking, Reading, Writing, Translation*.

Output: $\mathcal{L}$ is a learning path with an ordered sequence of N lessons, where each ℓ_i contains vocabularies V_i , scenario-specific dialogues D_i , and multimodal practice resources E_i .

3.2 Stage 1: Learning Path Scheduler

The Learning Path Scheduler decomposes the learner’s description S into pedagogically coherent sub-scenarios and fills the learning path with vocabulary, grammar, and cultural points.

Lesson Planning. To generate adaptive scenario-based curricula, the scheduler first constructs a lesson plan $\mathcal{P}$ through three steps: (1) **Learner Needs Analysis:** It summarizes the overall learning goal and corresponding advanced objectives for the target scenario. (2) **Scenario Modeling:** It decomposes the main scenario S into a logically ordered narrative flow of sub-scenarios, each corresponding to one lesson. (3) **Instructional Strategy Design:** It specifies teaching objectives and difficulty progression based on SMART principles (Wolf and Akkaraju, 2014), Bloom’s Taxonomy (Krathwohl, 2002), and Krashen’s Input Hypothesis (Krashen, 1982), assigning suitable difficulties to each lesson.

Pre-hoc Retrieval for Content Enrichment. We construct a Chinese Education Knowledge

Graph (CEKG) from Chinese proficiency standards, HSK syllabi, dictionaries, and cultural materials to provide structured pedagogical knowledge. Details of CCKG are provided in Appendix A. The scheduler retrieves lesson-relevant vocabulary, grammar, and cultural knowledge from CCKG to form C_{cekg} , which grounds the initial path generation:

$$\mathcal{L}_{init} = \text{LLM}(\mathcal{P}, C_{cekg}). \quad (2)$$

Post-hoc Retrieval for Personalization. The scheduler further refines $\mathcal{L}_{init}$ to mitigate hallucinations and improve personalization. Cognition alignment replaces out-of-level content with proficiency-appropriate alternatives, while preference alignment adjusts practice allocation according to learners’ skill preferences:

$$\mathcal{L} = \text{Align}(\mathcal{L}_{init}, \mathcal{U}, C_{cekg}). \quad (3)$$

Details are provided in Appendix B and Fig. 7.

3.3 Stage 2: Scenario Dialogue Generator

For each lesson ℓ_i in the learning path, the second-stage agent generates a core instructional dialogue D_i . The dialogue simulates authentic communication and naturally incorporates target knowledge points. Fig. 2 presents an example in which a customer discusses flavor preferences with a waiter in a Chinese restaurant. The agent further transforms each dialogue into multimodal scenario materials. It extracts key contextual elements from D_i , including roles, actions, objects, and environmental settings, to generate cartoon-style illustrations. It also

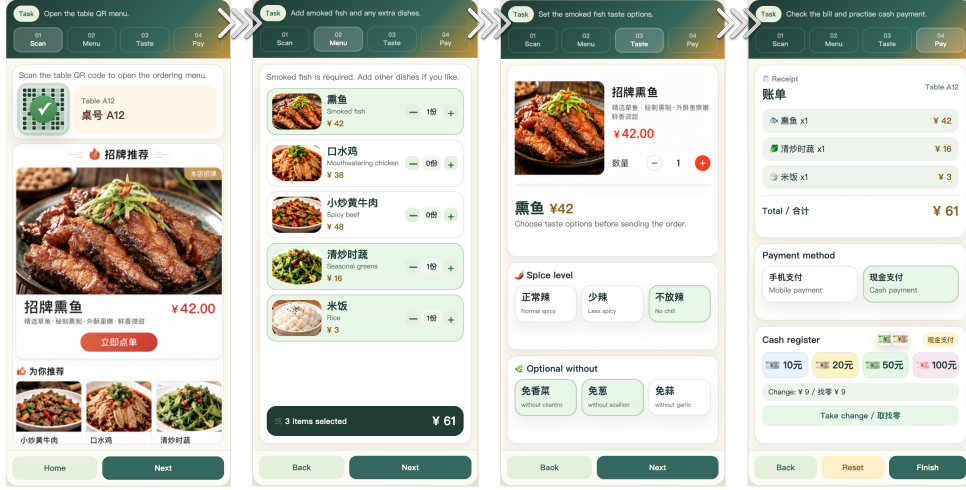


Figure 3: Example interactive simulation for Dining at a Chinese Restaurant

restructures the dialogue into a concise role-playing script for coherent video presentation. These materials help learners understand the communicative situation and prepare for interactive practice.

3.4 Stage 3: Interactive Practice Generator

The third stage augments the learning path $\mathcal{L}$ with multimodal practice activities E , including task-based interactive simulations and exercises.

Task-based Interactive Simulation. Instead of presenting static learning materials, interactive simulation enables learners to practice efficiently in an immersive task-based environment. The generation process consists of two steps: **(1) Interaction Flow Construction.** It first extracts the communicative goal, roles, scene objects, target language items, actions, and completion criteria to form a task specification. It then organizes them into a fine-grained interaction flow controlling task states, feedback, and progression within a sub-scenario. **(2) Simulation Instantiation.** The flow is instantiated as an editable HTML-based simulation with task prompts, interactive objects, choices, feedback, and progress controls. Correct actions advance the task, while inappropriate or incomplete actions trigger formative feedback and retries. This editable design allows instructors to refine the generated content according to specific pedagogical needs.

Dialogue-aware Multi-skill Exercises. Based on interactive simulations as context, another agent generates exercises covering *Listening*, *Speaking*, *Reading*, *Writing*, and *Translation*. To balance generation quality and efficiency, a hybrid strategy is applied. Rule-based programs generate structured tasks such as character writing and read-aloud

prompts. LLMs generate semantically demanding exercises such as comprehension, cloze, and translation questions. TTS and ASR services support audio-based listening and speaking practice.

4 Experiments and Evaluation

4.1 Experimental Setting

Dataset. To fairly evaluate the authenticity of generated scenarios, we construct a Chinese Textbook Dataset from the Official HSK Textbooks, covering six proficiency levels from HSK 1 to HSK 6. Each textbook is organized into topic-based units with scenario-based dialogues and vocabulary lists. After manual quality assessment, our dataset contains 120 distinct cases involving diverse scenarios.

Comparison Methods. To validate the generality of ImmersiveCFL, three typical LLM backbones are evaluated, including **Qwen3.6-Plus**, **DeepSeek-3.2**, and **GPT-5.4**. We compare with two baselines: **Direct Prompt**, where LLMs are directly prompted to generate learning paths, and **Chain-of-Thought (CoT) Prompt**, where LLMs reason using chain-of-thought prompting. For all baselines, **Qwen3.6-Plus** is used as the backbone due to its competitive ability and cost-effectiveness.

4.2 Quantitative Analysis

Vocabulary. We use Out-of-Vocabulary Rate (R_{OOV}) to measure generated words exceeding the target HSK level, where a small amount of above-level input is beneficial according to Krashen’s Input Hypothesis (Krashen, 1982), and Vocabulary Repetition Count ($\#Repeat$) to measure each word’s average exposure, following frequency-based second language acquisition (Ellis, 2013).

Model	R_{OOV} ($\downarrow$)	#Repeat ($\uparrow$)	DG ($\uparrow$)	Sim ($\uparrow$)	PPL ($\downarrow$)	SkillAlign. ($\uparrow$)
ImmersiveCFL						
with Qwen3.6-Plus	1.16%*	15.34*	0.4581*	74.24%*	9.91*	98.49%*
with DeepSeek-v3.2	0.73%	15.29	0.4637	73.18%	10.40	98.24%
with GPT-5.4	2.60%	18.90	0.5315	74.06%	9.61	98.52%
Direct Prompt	47.82%	5.39	0.0000	66.88%	10.99	72.05%
CoT Prompt	64.47%	4.58	0.0000	66.09%	11.60	69.29%

Table 1: Main experimental results. Arrows indicate the preferred direction for each metric. For R_{OOV} , values among 1%–3% are considered optimal according to Krashen’s Input Hypothesis (Krashen, 1982). * indicates a statistically significant difference ($p < 0.05$) between ImmersiveCFL and baselines.

As shown in Table 1, ImmersiveCFL maintains a low OOV rate across backbones, ranging from 0.73% to 2.60%, and repeats each word 15–18 times on average. Direct Prompt and CoT Prompt produce much higher OOV rates, 47.82% and 64.47%, with fewer repetitions, approximately 4–5 times per word, suggesting weaker difficulty control and vocabulary exposure.

Curriculum Content. We use the Difficulty Gradient of Learning Paths (DG), Adjacent Dialogue Similarity (Sim), and Perplexity of Generated Content (PPL) to evaluate curriculum content quality. DG measures the standard deviation of HSK levels across lessons to reflect the difficulty gradient of the learning path. Sim measures narrative continuity between adjacent dialogues using cosine similarity, and PPL reflects text fluency. As shown in Table 1, ImmersiveCFL achieves higher DG values, ranging from 0.45 to 0.53, while the baselines collapse to 0.00, indicating clearer progressive difficulty. It also obtains higher adjacent dialogue similarity, 73.18%–74.24%, compared with 66.09%–66.88% for baselines, suggesting better narrative continuity and relatively low PPL.

Personalization. To evaluate whether generated learning paths satisfy individual needs, Skill Alignment is defined as the cosine similarity between the multi-skill proportion distribution of generated exercises and the learner-specified skill preference distribution. As shown in Table 1, ImmersiveCFL achieves 98%–99% Skill Alignment, substantially outperforming the baselines at around 70%. More personalization and ablation results are provided in Appendix C and Appendix D.

4.3 User Study & Human Evaluation

To validate the pedagogical quality of ImmersiveCFL, we recruited 10 professional CFL teachers to evaluate it with the CoT baseline. We extracted 60 cases covering daily-life topics such as

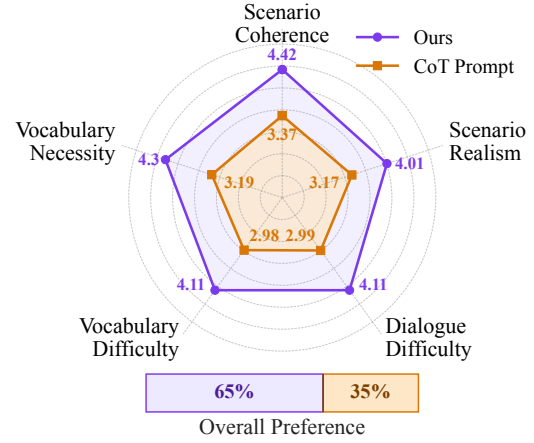


Figure 4: Human evaluation results.

ordering food, taking public transportation, or shopping. Each case was evaluated by at least 3 evaluators, resulting in 180 evaluation records. For each case, evaluators compared the anonymized curricula generated by two models side by side as *ResultA* and *ResultB*, and scored on 6 dimensions: scenario coherence, scenario realism, dialogue difficulty, vocabulary necessity, vocabulary difficulty, and preferred overall. Final results are derived from the average of all cases. As shown in Fig. 4, ImmersiveCFL outperforms the baseline by significant margins on all subjective metrics, and the number of evaluators who prefer ImmersiveCFL is nearly two times that of the baseline.

4.4 Platform Deployment

We have implemented ImmersiveCFL as a [public online platform](#) to demonstrate the generated immersive learning experience. Designed for CFL learners seeking scenario-specific practice and teachers preparing customized materials, the platform offers two entry points as illustrated in Fig. 5: *Generated Scenarios* and *Customize Scenarios*. In Generate Scenario Mode, users input a scenario prompt and configure preferences, and

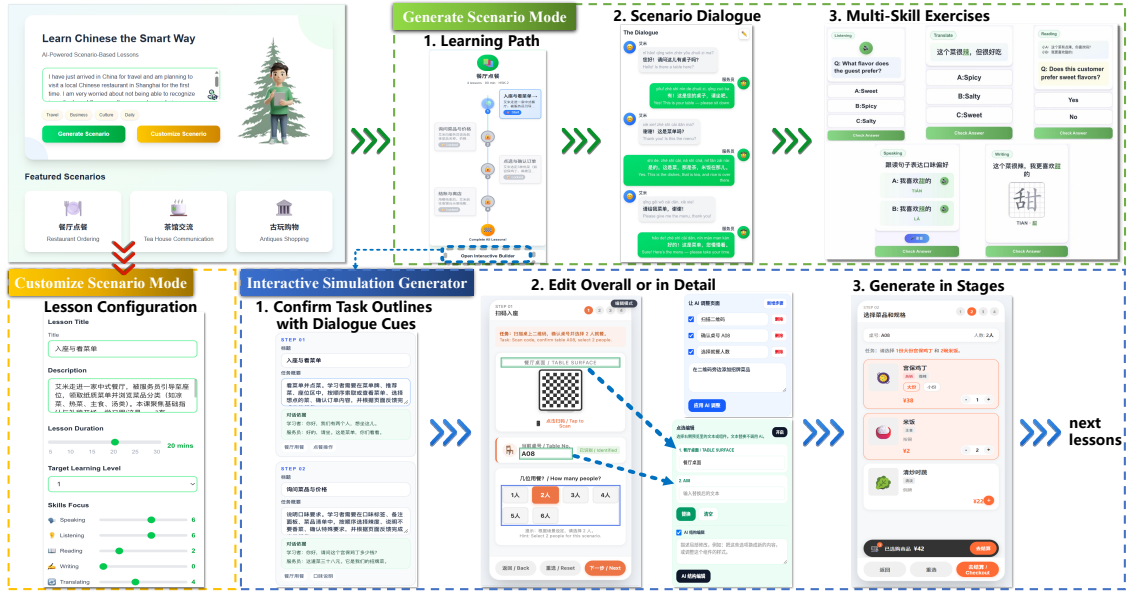


Figure 5: The functional pipeline of the deployed platform.

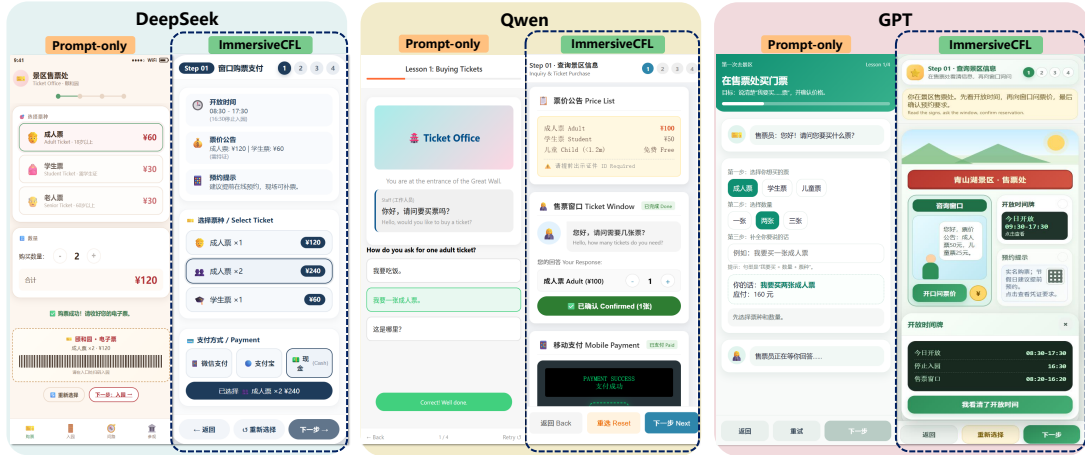


Figure 6: Interactive simulations comparison between ImmersiveCFL and prompt-only using three LLMs.

the system automatically generates an adaptive lesson sequence. For advanced learners or teachers, Customize Scenario Mode supports fine-grained intervention. The platform further provides interactive and immersive practices, including multi-skill exercises and interactive simulations.

4.5 Interactive Simulation Comparison

To intuitively demonstrate the quality of interactive simulations generated by ImmersiveCFL, we conduct a case-based comparison with the prompt-only baseline using three LLMs. As shown in Fig. 6, under the same scenario request, the prompt-only baseline tends to generate HTML pages resembling conventional forms or language-selection exercises, while ImmersiveCFL generates dialogue-grounded communicative tasks with contextual information,

manipulable objects, and immediate feedback. This comparison suggests that structured task grounding and staged generation help align communicative objectives and scenario context.

5 Conclusion

In this paper, we propose **ImmersiveCFL**, a multi-agent framework for generating immersive scenario-oriented CFL experiences. Given a learner-customized scenario and proficiency profile, ImmersiveCFL generates coherent learning paths, scenario-specific dialogues, interactive simulations, and multi-skill exercises with difficulty control. Experimental results demonstrate its potential for goal-driven Chinese language learning in terms of curriculum coherence, proficiency adaptation, and personalized practice.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62507015, 62537001).

References

- Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. M3-embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2318–2335, Bangkok, Thailand.
- Xianyu Chen, Jian Shen, Wei Xia, Jiarui Jin, Yakun Song, Weinan Zhang, Weiwen Liu, Menghui Zhu, Ruiming Tang, Kai Dong, et al. 2023. Set-to-sequence ranking-based concept-aware learning path recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5027–5035.
- Chinese Testing International. 2026. Hsk - hanyu shuiping kaoshi (chinese proficiency test). <https://www.chinesetest.cn/HSK>. Accessed on Apr. 30, 2026.
- Nick C Ellis. 2013. Frequency-based accounts of second language acquisition. In *The Routledge handbook of second language acquisition*, pages 193–210. Routledge.
- Edward Peter Errington. 2003. *Developing scenario-based learning: Practical insights for tertiary educators*. Dunmore Press.
- Robert M Gagne, Walter Wager, and Alicia Rojas. 1981. Planning and authoring computer-assisted instruction lessons. *Educational technology*, 21(9):17–26.
- Stephen Krashen. 1982. Principles and practice in second language acquisition.
- David R Krathwohl. 2002. A revision of bloom’s taxonomy: An overview. *Theory into practice*, 41(4):212–218.
- Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge university press.
- Qingyao Li, Wei Xia, Li’ang Yin, Jian Shen, Renting Rui, Weinan Zhang, Xianyu Chen, Ruiming Tang, and Yong Yu. 2023. Graph enhanced hierarchical reinforcement learning for goal-oriented learning path recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 1318–1327.
- Qi Liu, Shiwei Tong, Chuanren Liu, Hongke Zhao, En-hong Chen, Haiping Ma, and Shijin Wang. 2019. Exploiting cognitive structure for adaptive learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 627–635.
- Mike Long. 2014. *Second language acquisition and task-based language teaching*. John Wiley & Sons.
- Rui Lv, Qi Liu, Weibo Gao, Haotian Zhang, Junyu Lu, and Linbo Zhu. 2025. Genal: Generative agent for adaptive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 577–585.
- Ministry of Education of the People’s Republic of China and State Language Commission. 2021. [Chinese proficiency grading standards for international chinese language education \(gf 0025-2021\)](#). Technical report, Ministry of Education of the People’s Republic of China, Beijing, China.
- Morten Misfeldt. 2015. Scenario based education as a framework for understanding students engagement and learning in a project management simulation game. *Electronic journal of e-Learning*, 13(3):181–191.
- Amir Hossein Nabizadeh, José Paulo Leal, Hamed N Rafsanjani, and Rajiv Ratn Shah. 2020. Learning path personalization and recommendation methods: A survey of the state-of-the-art. *Expert Systems with Applications*, 159:113596.
- Rebecca L Oxford. 2006. Task-based language teaching and learning: An overview. *Asian EFL journal*, 8(3).
- Darwyn R Peachey and Gordon I McCalla. 1986. Using planning techniques in intelligent tutoring systems. *International Journal of Man-Machine Studies*, 24(1):77–98.
- M. S. Smith, S. Warnes, and A. Vanhoostenberghe. 2018. [Scenario-based learning](#). In J. P. Davies and N. Pachler, editors, *Teaching and Learning in Higher Education: Perspectives From UCL*, pages 144–156. UCL IOE Press, London, U.K. Accessed: Apr. 13, 2023. [Online].
- Julita Vassileva. 1992. Dynamic cal-courseware generation within an its-shell architecture. In *International Conference on Computer Assisted Learning*, pages 581–591. Springer.
- Julita Vassileva. 1995. Reactive instructional planning to support interacting teaching strategies. In *Proceedings of the 7-th World Conference on AI and Education*, pages 334–342. AACE Charlottesville, VA.
- Tianfu Wang, Yi Zhan, Jianxun Lian, Zhengyu Hu, Nicholas Jing Yuan, Qi Zhang, Xing Xie, and Hui Xiong. 2025. Llm-powered multi-agent framework for goal-oriented learning in intelligent tutoring system. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 510–519.
- Jane Willis. 2021. *A framework for task-based learning*. Intrinsic Books Ltd.

- Alexander Wolf and Shylaja Akkaraju. 2014. Teaching evolution: From smart objectives to threshold experience. *Journal of Effective Teaching*, 14(2):35–48.
- Xueqiao Zhang, Chao Zhang, Jianwen Sun, Jun Xiao, Yi Yang, and Yawei Luo. 2025. Eduplanner: Llm-based multi-agent systems for customized and intelligent instructional design. *IEEE Transactions on Learning Technologies*.
- Yuwen Zhou, Changqin Huang, Qintai Hu, Jia Zhu, and Yong Tang. 2018. Personalized learning full-path recommendation model based on lstm neural networks. *Information sciences*, 444:135–152.

A Chinese Education Knowledge Graph

We crafted a **Chinese Education Knowledge Graph (CEKG)** as the knowledge foundation of ImmersiveCFL. The data are collected from multiple sources, including Chinese Proficiency Grading Standards for International Chinese Language Education (Standards) (Ministry of Education of the People’s Republic of China and State Language Commission, 2021), Hanyu Shuiping Kaoshi (HSK) syllabi (Chinese Testing International, 2026), dictionaries, and cultural materials. We first clean and normalize the raw data, then use entity recognition and relation extraction to construct candidate triples, and finally apply cross-source entity alignment and manual checking to reduce duplicates and improve data quality.

Table 2 summarizes the statistics of CEEKG. Different from a simple vocabulary list, CEEKG organizes Chinese learning resources as a structured semantic network. The CEEKG contains several types of entities, such as Hanzi, vocabulary, grammar points, HSK levels, learner errors, and cultural points, each entity is stored with teaching-oriented attributes. The relations in CEEKG describe how different knowledge items are connected, including *synonym*, *antonym*, *co-occurrence*, word formation, *learning dependency*, *entity difficulty*, and *cultural association*. Beyond storing isolated knowledge entities, CEEKG supports relation-aware retrieval, precise difficulty control, and cultural enrichment, providing a more reliable knowledge foundation for LLM-based generation.

Resource	Count	Data Source
Entity		
Hanzi	20,841	Standards, HSK Syllabi, Stroke Order
Vocabulary	267,853	Standards, HSK Syllabi, Dictionaries
Idioms	56,752	Idiom Dictionaries
Level Types	20	Standards, HSK Syllabi
Grammar Points	1,167	Standards, HSK Syllabi
Learner Errors	602	Chinese Writing Error Resources
Cultural Points	591	Cultural Teaching Reference Data
Relation		
Synonym	1,152	Chinese Synonym Dictionary
Antonym	5,375	Online Chinese Dictionaries
Co-occurrence	43,363	Baidu Baike Corpus
Entity-Level	31,744	Standards, HSK Syllabi
Word Formation	557,263	Character-Word Composition Data
Cultural Association	9,613	Cultural Teaching Resources
Learning Dependency	435	Radical-Based Hanzi Relations

Table 2: Statistics of CEEKG.

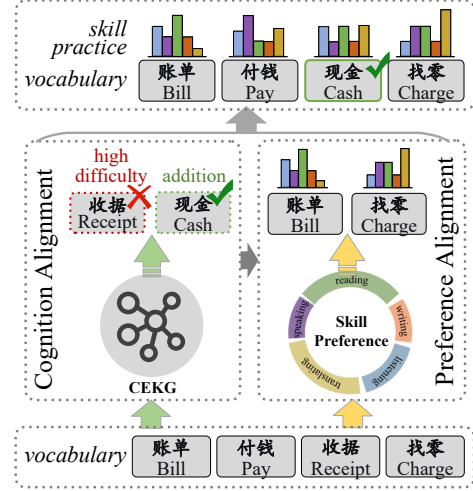


Figure 7: Example of Post-hoc Retrieval for Cognition and Preference Alignment.

B Post-hoc Retrieval Alignment

As shown in Fig. 7, the scheduler refines the learning path through cognition and preference alignment. For cognition alignment, out-of-level vocabulary is replaced by semantically similar words whose HSK level does not exceed the lesson target level. Word embeddings from **bge-large-zh-v1.5** (Chen et al., 2024) are used to select replacements with high semantic cosine similarity. For preference alignment, the scheduler allocates learning time under the duration constraint T . Harder words are assigned larger learning weights, and the final exercise time is further distributed across *Listening*, *Speaking*, *Reading*, *Writing*, *Translation* according to both pedagogical skill weights and the learner’s preference vector $\mathbf{w}_{skill}$.

C Personalization Analysis

As shown in Fig. 9, the left figure reports the average sentence length of generated dialogues across learner proficiency levels, and the right figure reports the average number of exercises under different learning durations. ImmersiveCFL captures differences in learner proficiency by generating proficiency-sensitive dialogues. It also increases exercise counts according to lesson duration. By contrast, the baselines tend to generate shorter dialogues and fewer exercises, showing weaker adaptation to personalized time requirements.

D Ablation Study

We conduct ablation studies on two components of the Stage 1 Learning Path Scheduler: Lesson



Figure 8: Example Scenario-Based Instructional Content for Taking the Shanghai Metro

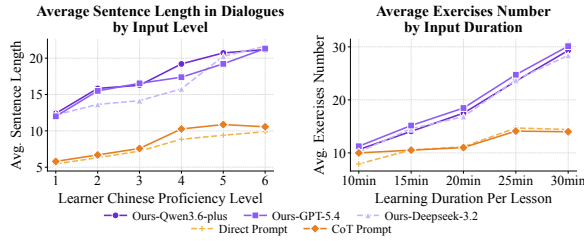


Figure 9: Personalization analysis across learner proficiency levels and lesson durations.

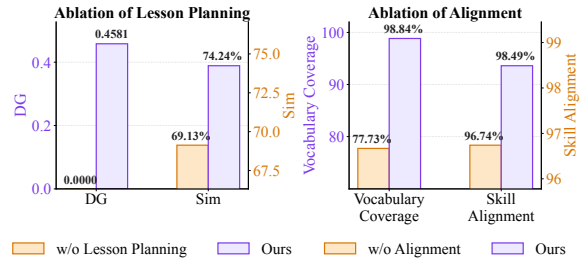


Figure 10: Ablation study results.

Planning and the Alignment module, which performs cognition and preference alignment. Removing Lesson Planning eliminates the decomposition of learner-described scenarios into progressive sub-scenarios and weakens the reconstruction of a logically ordered narrative flow. Removing Alignment causes the model to generate content without learner-specific cognition and preference alignment.

As shown in Fig. 10, without Lesson Planning, the Difficulty Gradient collapses to 0.00 and Adjacent Dialogue Similarity decreases, indicating that the learning path lacks progressive structure and narrative coherence. Without Alignment, Vocabulary Coverage, defined as $1 - R_{OOV}$, and Skill Alignment decrease substantially. These results show that Lesson Planning supports pedagogical progression, while Alignment improves vocabulary difficulty control and skill preference alignment.

E Case Study

As shown in Fig. 8, we present another generated example for the scenario of *Taking the Shanghai Metro*. The scenario is decomposed into four ordered sub-scenarios, including station search and

ticket purchase, asking for directions, line transfer, and arrival and exit. Each lesson contains multimedia dialogue, vocabulary, exercises and interactive simulation. Taking the first lesson as an example, we create dialogue text, instruction images and videos about finding the metro entrance and purchasing a ticket to the Bund, these dialogues include a range of pedagogically appropriate vocabulary. The lesson follows a suite of multimodal practice, including multi-skill exercises and HTML-based interactive simulations, where learners can practice the various linguistic skills for determining directions or transferring lines. This case demonstrates that integrating a learning path, dialogue, and interactive practice into a unified scenario context enables learners to practice Chinese through immersive and goal-driven interaction.