Received: 01 Dec. 2025 | Revised: 05 Dec. 2025 | Accepted: 08 Dec. 2025 | Published online: 25 Dec. 2025

RESEARCH ARTICLE

Contemporary Education and Teaching Research
2025, Vol. 6 (12)532-554
DOI: 10.61360/BoniCETR252019451201

AI-Based Virtual Environments for Chinese EFL Learners: A Three-Layer Developmental Model with Pilot Evidence

Jun Zou[1,2] & Beini Ma[1,*]
[1]*School of Foreign Languages, Shaoxing University, China*
[2]*School of International Education, Shaoxing University, China*

Abstract: The rapid advancement of artificial intelligence (AI) and immersive technologies such as virtual reality (VR), augmented reality (AR), mixed reality (MR), and large language models (LLMs) is reshaping foreign language education worldwide. In the Chinese EFL context, however, traditional classroom-based instruction still suffers from limited authentic input, constrained opportunities for interaction, delayed and non-individualized feedback, and insufficient exposure to pragmatic and intercultural experiences. Drawing on Second Language Acquisition (SLA) theories, constructivism, and situated learning, this study proposes a three-layer developmental model that explains how AI-based virtual environments (AI-VEs) can support Chinese EFL learners' linguistic, communicative, and intercultural development. At the outer layer, diversified immersive scenarios provide ecologically valid contexts; at the middle layer, a recursive Input–Interaction–Output–Reflection (IIOR) cycle captures core learning mechanisms; at the inner layer, learners' competencies develop from lexical and formulaic knowledge toward communicative and intercultural competence. The conceptual model is illustrated through a small-scale pilot study involving twelve Chinese undergraduates who completed two AI-VE tasks. Mixed-methods analyses of oral production, interaction logs, questionnaires, and interviews indicate gains in lexical sophistication, increased negotiation of meaning, greater use of formulaic sequences, higher willingness to communicate, and enhanced intercultural sensitivity. These findings offer initial empirical support for the proposed model and suggest that AI-VEs can function as powerful mediational tools for advancing EFL education and educational equity in China. Implications for curriculum design, teacher professional development, and future research are discussed.

Keywords: AI virtual environments, language development, EFL learners, SLA, intelligent feedback

1. Introduction

Recent advances in artificial intelligence, including large language models (LLMs), natural language processing (NLP), automated speech recognition (ASR), and immersive technologies such as VR, AR, and MR, are rapidly transforming the landscape of language education. For Chinese EFL learners, traditional instructional ecologies—characterized by classroom-bound teaching, textbook-driven materials, and limited real communicative exposure—have long constrained opportunities for authentic input, meaningful interaction, and contextually appropriate language use. Despite national efforts to elevate English education, learners frequently demonstrate imbalances between linguistic knowledge and real-world communicative performance, reflecting systemic limitations related to insufficient input richness, restricted interactional possibilities, delayed or generalized feedback, and minimal exposure to pragmatic or intercultural experiences.

Corresponding Author: Beini Ma
School of Foreign Languages, Shaoxing University, China

AI-based virtual environments (AI-VEs) offer a compelling response to these challenges by constructing immersive, contextually rich, and socially situated learning spaces that approximate—and at times enhance—the complexity of real communicative encounters. Through high-fidelity simulations, multimodal semiotic resources, and adaptive conversational agents, AI-VEs enable learners to engage in repeated, low-stakes communicative practice; negotiate meaning across diverse scenarios; receive immediate, fine-grained, and personalized feedback; and develop linguistic, pragmatic, and intercultural competencies through continuous performance analytics. These affordances operationalize, in technologically augmented form, the core mechanisms theorized to drive second language acquisition, including comprehensible input, interactional work, pushed output, and metacognitive reflection.

However, the existing literature on AI and immersive technologies in language learning remains fragmented. Research tends to prioritize short-term performance effects, offering limited insight into longitudinal developmental trajectories. Few studies provide theoretically integrated models that explicitly connect SLA constructs with the technological affordances of AI-VEs, and mechanism-oriented accounts of how AI-VEs mediate linguistic, pragmatic, and intercultural growth remain underdeveloped. Moreover, studies addressing the specific sociocultural and educational needs of Chinese EFL learners within AI-VE contexts are scarce, restricting the contextual validity and applicability of existing findings.

In response, the present study proposes a theoretically grounded, mechanism-driven developmental model that synthesizes SLA theory with the pedagogical and technological affordances of AI-VEs. Specifically, it aims to elucidate how AI-VEs facilitate language development among Chinese EFL learners; analyze the roles of input, interaction, output, and reflection within these environments; articulate a three-layer developmental framework capturing linguistic, communicative, and intercultural progression; and identify the pedagogical and social implications of integrating AI-VEs into contemporary language education.

Accordingly, this study pursues two main aims: (1) to develop a theoretically grounded, mechanism-oriented model of how AI-based virtual environments can support Chinese EFL learners' language development; and (2) to provide initial empirical illustrations of this model through a small-scale pilot study.

To address these aims, the study is guided by the following research questions:

RQ1: How can AI-based virtual environments be conceptualized as supporting Chinese EFL learners' linguistic, communicative, and intercultural development?

RQ2: Through which key mechanisms do AI-based virtual environments facilitate the Input–Interaction–Output–Reflection (IIOR) cycle?

RQ3: To what extent do findings from a pilot implementation provide initial empirical support for the proposed three-layer developmental model?

2. Literature Review

This section synthesizes a range of theoretical, technological, and empirical strands that collectively inform the present study. Drawing on foundational constructs in Second Language Acquisition (SLA), emerging research on AI-driven and immersive technologies, and recent developments in the Chinese EFL landscape, it establishes the conceptual architecture necessary for theorizing the role of AI-based virtual environments (AI-VEs) in mediating multidimensional language development. The review not only identifies converging insights but also foregrounds persistent conceptual and methodological lacunae that the present study seeks to address.

2.1 SLA theories

SLA research has long provided the conceptual scaffolding for understanding how language learning unfolds across cognitive, social, and interactional dimensions. These theories are particularly salient for explicating how AI-VEs may catalyze interlanguage development through technologically mediated forms of input, interaction, output, and reflection.

2.1.1 Krashen's input hypothesis and multimodal comprehensible input

Krashen's Input Hypothesis (1985) posits that acquisition is driven by exposure to comprehensible input that is slightly beyond learners' current competence (i+1). In many Chinese EFL classrooms, such input is limited by textbook-based materials and relatively uniform syllabi. AI-based virtual environments, by contrast, can deliver context-rich, multimodal, and adaptively scaffolded input, thereby operationalizing comprehensible input in technologically enhanced ways.

2.1.2 Long's interaction hypothesis and negotiation of meaning

Long's Interaction Hypothesis (1996) foregrounds the role of interactional adjustments—such as clarification requests, confirmation checks, and recasts—in facilitating acquisition by increasing the salience of linguistic features (Gass, 1997). AI-VEs, powered by natural language understanding (NLU) and speech processing technologies, are capable of replicating and even amplifying these interactional contingencies. Virtual interlocutors can engage learners in multi-turn dialogue sequences, detect communicative breakdowns, and initiate negotiation-of-meaning moves with precision and immediacy. Empirical studies (Peterson, 2012) confirm that virtual environments can foster the same collaborative interactional dynamics that underpin human–human SLA, thereby providing a fertile ground for interlanguage restructuring (Long, 1996).

2.1.3 Swain's output hypothesis and pushed output

Swain (1995) argues that linguistic production is not merely a manifestation of acquired knowledge but a driving force in acquisition, especially when learners are "pushed" to produce more accurate, complex, and coherent language. AI-VEs embed learners in task-based communicative situations—ranging from service encounters to academic exchanges—that demand pragmatically appropriate, syntactically encoded, and semantically precise output. Learners must articulate meaning, monitor their performance, reformulate utterances based on feedback, and draw on strategic and pragmatic resources. Through these cognitively taxing conditions, AI-VEs instantiate the metalinguistic and syntactic processing that Swain identifies as central to development (Swain, 1995).

2.1.4 The CAF framework

The CAF triad has emerged as a robust empirical framework for assessing linguistic performance (Housen, Kuiken & Vedder, 2012). AI-VEs are uniquely suited to support CAF-oriented research and pedagogy because they generate fine-grained, longitudinal, and multimodal learner data, enabling analyses of: developmental complexity (lexical diversity, syntactic elaboration), accuracy trajectories (morphosyntactic precision, pragmatic appropriateness), fluency patterns (temporal measures, repair sequences).

Automated learning analytics embedded in AI-VEs allow for sustained monitoring of CAF indicators, offering insights that surpass the granularity feasible in traditional classroom observations (Larsen-Freeman, 2006).

2.1.5 Sociocultural theory and situated learning

From a sociocultural perspective (Vygotsky, 1978), learning is mediated through socially situated interaction (Vygotsky, 1978), while Situated Learning Theory (Lave & Wenger, 1991) emphasizes participation within authentic communities of practice (Lave & Wenger, 1991). AI-VEs provide simulated social ecologies wherein learners can enact roles, engage in joint activity, and participate in communicative events reflective of real-world norms and expectations. These environments support scaffolded participation within a safe, adaptive, and socially meaningful space, aligning closely with sociocultural notions of mediated learning and apprenticeship.

2.2 Constructivist and situated learning

The integration of AI, VR, AR, and other immersive technologies has introduced new possibilities for language learning that transcend the limitations of text-based or classroom-bound instruction.

VR environments provide uniquely immersive learning experiences by creating a strong sense of presence, or the perceptual illusion of "being there," while enabling sensorimotor immersion that allows learners to interact with virtual objects in an embodied manner (Lan, 2020). They also offer a high degree of contextual authenticity, supporting task performance in realistic settings, and facilitate interactional reciprocity through two-way communication with virtual agents or avatars (Graesser, 2016). Moreover, VR environments promote embodiment by allowing learners to inhabit virtual identities, thereby deepening engagement and enhancing the authenticity of communicative experiences.

Empirical evidence (Yeh & Lan, 2018) demonstrates that such immersion reduces anxiety, enhances willingness to communicate, and improves pragmatic competence by enabling learners to engage spontaneously without fear of social evaluation (MacIntyre et al., 1998).

Advances in AI-driven NLP and speech technologies—including ASR, TTS, conversational agents, and LLM-powered dialogue systems—enable virtual environments to approximate human-like interactional responsiveness. These technologies allow AI-VEs to: assess learner speech in real time, diagnose phonological, morphosyntactic, and pragmatic deviations, deliver context-sensitive recasts and feedback, maintain coherent multi-turn discourse.

The result is a dynamic communicative environment in which feedback loops are immediate, adaptive, and continuous.

AI-driven analytics systems identify: recurrent error patterns, developmental trajectories, fluency rhythms, lexical distribution patterns, task completion behaviors. Learners gain access to personalized dashboards that visualize performance trends, promote metacognitive reflection, and orient subsequent learning efforts. Studies (Li et al., 2020) affirm that AI-generated feedback can rival human feedback in precision and pedagogical effectiveness (Luckin, 2017), particularly in resource-limited educational contexts.

2.3 AI and immersive technologies in language education

As China accelerates its digital transformation in education, research on AI-mediated language learning is expanding, though often unevenly distributed across regions and educational levels (Wang & Vásquez, 2014).

Chinese studies indicate that integrating multimedia and VR technologies can: enhance engagement and motivation, support vocabulary and pronunciation development, foster meaningful interaction. However, much of this research remains confined to short-term interventions, with limited theorization or broader generalizability (Li, Link & Hegelheimer, 2015).

Research on AWE systems, AI-based speaking assessment, and recommendation algorithms suggests that AI tools can improve: grammatical accuracy, lexical sophistication, argumentation coherence, oral fluency.

Yet few studies examine how such tools contribute to long-term developmental trajectories, a critical gap for understanding sustained SLA.

VR/AR-based studies in China show improvements in: pragmatic sensitivity (Zhao, 2005), WTC (willingness to communicate), anxiety reduction, communicative confidence (Radianti et al., 2020).

Nevertheless, the link between immersive environments and underlying SLA mechanisms remains underexplored, leaving theoretical questions unresolved (Pellas, 2014).

2.4 Research gaps

Although research on AI and immersive technologies is expanding rapidly, several important gaps remain. There is a lack of theoretically integrated models that connect SLA mechanisms with the specific affordances of AI-VR environments, and empirical studies have paid insufficient attention to the sociocultural and institutional needs of Chinese EFL learners. In addition, our understanding of the longitudinal developmental trajectories supported by AI-driven scaffolding is still limited, and pragmatic as well as intercultural development in virtual environments remains under-theorized (Yeh & Lan, 2018). Furthermore, existing research rarely offers multi-layered frameworks that explain how contextual ecosystems, learning cycles, and developmental processes interact to shape language learning outcomes.

This study seeks to address these gaps by proposing a comprehensive, mechanism-oriented, theoretically grounded model for understanding how AI-VEs can support Chinese EFL learners' language development across linguistic, communicative, and intercultural dimensions.

3. Theoretical Framework

This section presents the theoretical foundations that underpin the proposed model for language development in AI-based virtual environments (AI-VEs). Building on established theories in Second Language Acquisition (SLA), constructivist perspectives, and situated learning, it explains how AI-VEs operationalize these theoretical principles through immersive, interactive, and adaptive mechanisms. Furthermore, this section introduces the Input–Interaction–Output–Reflection Cycle, which acts as the central learning mechanism integrating cognitive, social, and technological elements.

3.1 Second language acquisition theories supporting the model

Second Language Acquisition (SLA) theories offer a broad yet integrated account of how linguistic, cognitive, and interactional processes collectively shape the trajectory of language development. These theories articulate the mechanisms by which learners internalize linguistic input, engage in interactional work, and restructure their evolving interlanguages. As such, they provide a conceptual foundation for understanding how AI-based virtual environments (AI-VEs)—with their multimodal, adaptive, and socially simulated affordances—can catalyze and extend established SLA processes in ways traditional pedagogical contexts cannot easily achieve.

3.1.1 Comprehensible input in AI-based virtual environments

Krashen's Input Hypothesis posits that acquisition is driven by exposure to comprehensible input located just beyond the learner's current competence (i+1). Traditional EFL classrooms, constrained by curricular uniformity, limited exposure conditions, and teacher-dependent variability, struggle to deliver input that is simultaneously authentic, context-rich, and adaptively scaffolded.

AI-VEs, however, fundamentally transform the ecology of input. Through multimodal semiotic resources—visual, auditory, gestural, and spatial—they anchor language in perceptually rich contexts that enhance traceability and grounding. Additionally, dynamic environmental cues, ranging from spatial layouts to paralinguistic signals, endow linguistic forms with immediate situational relevance. Crucially, AI-VEs deploy adaptive speech and text generation calibrated to continuously updated learner profiles, enabling the system to maintain input within the learner's zone of proximal development (VanPatten, 2015). Situated dialogues, embedded in meaningful communicative episodes, further contextualize linguistic forms within goal-oriented activity. In aggregate, these affordances instantiate a technologically amplified form of comprehensible input that is consistent with multimodal learning theory and far exceeds the representational capacity of conventional instructional settings.

3.1.2 Interaction and negotiation of meaning

Long's Interaction Hypothesis underscores the pivotal role of interaction-induced modifications in driving acquisition. Clarification requests, confirmation checks, recasts, and other negotiation moves enhance the salience of linguistic forms by directing learner attention to mismatches between intention and output (Schmidt, 1990).

AI-VEs have the unique capacity to reproduce interactional contingencies with remarkable precision and scalability. Leveraging advances in natural language processing, speech recognition, and conversational AI, virtual interlocutors can detect communicative breakdowns, initiate negotiation of meaning, issue context-sensitive prompts, and stimulate elaborated learner responses, thereby closely mirroring the dynamics of human interaction in second language acquisition.

Importantly, these moves are not bound by the temporal or interpersonal constraints typical of human-human dialogue. Instead, AI-VEs generate high-density interactional episodes where learners face repeated opportunities to refine utterances, monitor their interlanguage, and attend to form–meaning discrepancies. As learners engage in sustained negotiation sequences, cognitive processing deepens, promoting durable interlanguage restructuring.

3.1.3 Pushed output and language production

Swain's Output Hypothesis highlights the epistemic function of language production: learners produce language not merely as a communicative act but as a cognitive tool for hypothesis testing, self-monitoring, and noticing linguistic gaps.

AI-VEs operationalize this principle through task-based communicative episodes that require learners to plan, articulate, and refine utterances under conditions that demand accuracy, complexity, and pragmatic appropriateness. Whether negotiating solutions in collaborative simulations, participating in institutional interactions, or engaging in role-play dialogues, learners must employ strategic competence, invoke higher-order syntactic structures, and adapt their language to evolving discourse demands.

These contexts create sustained pressure for pushed output, stimulating metalinguistic reflection and strengthening the form–function mappings underlying advanced interlanguage development.

3.1.4 The CAF perspective and developmental monitoring

The Complexity–Accuracy–Fluency (CAF) triad has become central to empirical research on performance-based language development. AI-VEs align seamlessly with CAF-oriented inquiry because they capture high-resolution data across the temporal and linguistic dimensions of learner behavior.

Within AI-VEs, algorithms can systematically monitor key dimensions of learner performance, including complexity—reflected in syntactic elaboration and lexical sophistication—accuracy, which captures developmental patterns of morphosyntactic stabilization, and fluency, assessed through indicators such as speech rate, hesitation phenomena, and repair trajectories.

Such analytics surpass the observational capacity of traditional classrooms, enabling both real-time diagnostic feedback and longitudinal developmental modeling. The result is an operationalized, system-level form of CAF measurement embedded within everyday communicative activity.

3.2 Constructivism and learner agency

Constructivist theory conceptualizes learning as an active, integrative process in which learners construct knowledge through engagement with their environment (Cobb, 1994). AI-VEs embody this epistemological view by positioning learners as agents within goal-directed, inquiry-rich environments.

Within AI-VEs, learners navigate complex scenarios, make decisions with communicative consequences, experiment with a range of linguistic options, manipulate virtual objects and interlocutors, and co-construct meaning with virtual agents, thereby engaging actively in situated and dynamic language-use processes.

These experiences foster cognitive engagement, deepen conceptual grounding, and transform learners from passive recipients of prepackaged content into co-creators of situational meaning. The autonomy and agency supported by AI-VEs align with contemporary understandings of learner-driven exploration and experiential learning.

3.3 Situated learning and authentic contexts

Situated Learning Theory (Lave & Wenger, 1991) posits that learning arises through legitimate participation in socially meaningful practices. From this perspective, language learning requires not only exposure to forms but participation in socioculturally situated activity systems.

AI-VEs enable simulated authenticity by orchestrating communicative events that mirror real-world demands. Learners inhabit contextualized identities—such as customers, applicants, presenters, or mediators—and engage in role-based practices that reflect the pragmatic and cultural expectations of real communities. Through these situated interactions, learners acquire pragmatic norms, develop socio-interactional awareness, and participate in culturally grounded communicative behaviors. This alignment between situated practice and virtual experience substantially enhances ecological validity.

3.4 The Input–Interaction–Output–Reflection cycle

The mechanisms discussed above converge in a recursive, self-reinforcing cycle central to AI-VE–mediated development: the Input–Interaction–Output–Reflection (IIOR) Cycle. AI-VEs support each stage individually while integrating them into a dynamic system of ongoing development.

Learners receive multimodal, context-rich input via speech, visual cues, environmental affordances, and paralinguistic signals. Unlike textbook-bound representations, such input reflects authentic communicative conditions and promotes both immersion and comprehension.

Responsive avatars facilitate negotiation of meaning, register variation, and discourse-level engagement. Interaction becomes a site for noticing, alignment, and hypothesis testing as learners encounter shifting communicative demands.

Learners participate in performance-based tasks requiring meaningful language production. Output processes strengthen syntactic encoding, semantic precision, and strategic competence.

AI-generated analytics provide immediate feedback, enabling learners to engage in metacognitive evaluation, error analysis, and self-regulated adaptation. Reflection enhances consolidation and guides subsequent learning cycles. Together, these stages form a recursive developmental engine, activating interconnected SLA processes through immersive technological mediation.

Figure 1 illustrates the IIOR cycle as a dynamic, iterative mechanism whereby multimodal input, interactional work, communicative production, and reflective processes interlock to generate sustained development.

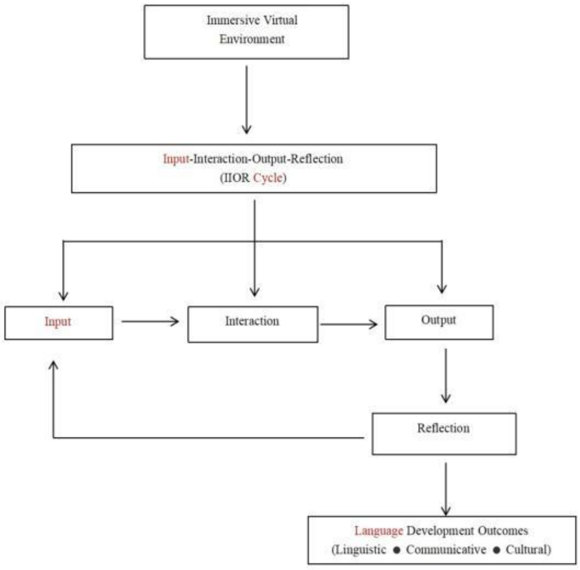

Figure 1 The IIOR Cycle

3.5 Integration of SLA, constructivism, and situated learning

AI-VEs serve as a convergence point for multiple theoretical perspectives: SLA elucidates the cognitive and interactional mechanisms underlying language acquisition, constructivism highlights the learner’s agency, exploration, and active construction of knowledge, and situated learning situates communication within authentic social practices. Together, these perspectives underscore how AI-VEs integrate cognitive, experiential, and contextual dimensions to support holistic language development. By merging these theoretical lenses, AI-VEs create immersive learning ecologies where learners engage in meaning-making, negotiate identities, and receive continuous feedback across cognitively and socially rich contexts.

This theoretical framework demonstrates that AI-based virtual environments not only align with but extend core SLA principles, constructivist epistemologies, and situated learning perspectives. Through immersive contexts, adaptive scaffolding, and real-time analytic feedback, AI-VEs activate the full spectrum of processes implicated in language development.

The Input–Interaction–Output–Reflection cycle and the subsequent multi-layer developmental model emerge naturally from these integrated theoretical foundations.

4. Methodology

4.1 Participants

Twelve Chinese EFL undergraduates (7 females, 5 males; aged 18–21) from a comprehensive university in eastern China voluntarily participated in this pilot study. Participants were recruited through an open call in an English elective course. Prior to the study, all learners completed a background questionnaire assessing their linguistic history, technology use, and prior exposure to immersive learning environments. All learners reported intermediate English proficiency (self-rated CEFR B1–B2) and had no previous experience with virtual reality or AI-based virtual environments.

Participation was voluntary, and informed consent was obtained from all individuals. Learners were assured that their participation—or decision not to participate—would not influence their course grades. All data were anonymized prior to analysis to ensure confidentiality and ethical compliance.

4.2 AI-VE platform and tasks

The study employed a custom-built AI-based Virtual Environment (AI-VE) that integrates adaptive speech technology, natural language understanding, multimodal input channels, and real-time learning analytics. The platform includes avatar-mediated communication, 3D interactive environments, and dynamic scaffolding mechanisms such as paraphrasing, speech-rate control, and visual highlighting.

Learners completed two task-based scenarios designed to elicit spontaneous language production in ecologically valid contexts:

(1) Task 1: Airport Check-in Interaction

Learners engaged in a simulated check-in conversation with an AI avatar functioning as an airline agent. The scenario required them to: provide personal information, ask clarifying questions, respond to unexpected complications (e.g., overweight baggage, seat changes). This task was designed to elicit lexical production, negotiation of meaning, and formulaic service-encounter expressions.

(2) Task 2: Cross-cultural Group Meeting

Learners participated in a virtual meeting involving a misunderstanding with an international team member. The task required: interpreting culturally nuanced utterances, expressing disagreement politely, resolving intercultural misunderstandings. This task targeted pragmatic competence, discourse management, and intercultural sensitivity.

Each task lasted approximately 25 minutes, and all interactions were logged automatically by the system for subsequent analysis.

4.3 Data collection procedures

Data were collected through four complementary sources to ensure methodological triangulation.

(1) Oral Production Data

Learners' speech during both tasks was audio-recorded and automatically transcribed using the AI-VE's built-in ASR system, then manually corrected for accuracy. These transcripts were used for: CAF (Complexity–Accuracy–Fluency) analysis, lexical sophistication calculation, formulaic sequence identification.

(2) Interaction Logs

The AI-VE platform logged: all learner–avatar turns, system-generated prompts, recasts, clarification requests, hesitation durations, speech-rate adaptations. Interaction logs were used to code negotiation of meaning episodes according to Long's (1996) taxonomy.

(3) Questionnaires

Learners completed three validated scales immediately before and after the tasks: Willingness to Communicate Scale (WTC), Perceived Presence Scale (VR immersion measure), Intercultural Sensitivity Scale (ISS)

Responses were rated on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree). Reliability analyses showed acceptable internal consistency (Cronbach's α = .82–.89).

(4) Semi-structured Interviews

At the end of the study, all learners participated in 12–15 minute interviews conducted in Mandarin to allow for richer expression. Interviews explored: perceptions of multimodal input, interactional experiences, emotional responses, changes in intercultural awareness.

All interviews were recorded, transcribed verbatim, and translated for analysis.

4.4 Data analysis

A mixed-methods analysis procedure was adopted to interpret quantitative and qualitative data in a complementary manner.

4.4.1 CAF analysis

CAF indices were computed using established procedures (Housen, Kuiken & Vedder, 2012): Complexity, Mean Length of T-unit (MLT), Clauses per T-unit. Lexical sophistication, assessed via the proportion of mid-/low-frequency lexical items (Zipf score < 3.5); Accuracy, Error-Free T-unit Ratio (EF-T), Error Density (errors per 100 words), Errors included grammatical, lexical, and morphological mistakes; Fluency, Speech Rate (words per minute), Mean Length of Pauses (>0.3s threshold), Self-repair Frequency, coded following Kormos (2006).

4.4.2 Interaction analysis

Negotiation of meaning episodes were identified and coded according to Long (1996) and modified frameworks by Lyster & Ranta (1997), categorizing: clarification requests, confirmation checks, recasts, comprehension checks. Learner uptake was coded as successful or unsuccessful depending on whether learners produced a modified output.

4.4.3 Questionnaire analysis

Pre/post differences on WTC, Presence, and ISS scores were analyzed descriptively, given the small sample size. Reliability was checked using Cronbach's alpha. Changes were interpreted holistically, supplemented by qualitative data.

4.4.4 Interview analysis

Interview transcripts were analyzed using thematic analysis: Open coding to identify preliminary categories, Axial coding to cluster categories into themes, Selective coding to align emergent themes with the six mechanisms coded in Section 4 Themes included: multimodal scaffolding, affective support, negotiation sensitivity, intercultural reflection. These qualitative insights triangulated with the quantitative results.

5. Mechanisms of Language Development in AI-based Virtual Environments

AI-based virtual environments (AI-VEs) support second language development through a constellation of mutually reinforcing cognitive, social, interactional, and affective mechanisms. These mechanisms both instantiate and extend core constructs in Second Language Acquisition (SLA), while leveraging technological affordances that are unique to AI-driven, immersive systems. In what follows, six interrelated mechanisms are elaborated: (1) the enhancement of comprehensible and multimodal input, (2) the deepening of interaction and negotiation of meaning, (3) the promotion of routinization and automatization, (4) the creation of personalized and adaptive learning pathways, (5) emotional engagement and motivational support, and (6) the cultivation of intercultural competence.

5.1 Enhancement of comprehensible and multimodal input

A primary mechanism through which AI-VEs foster language development lies in their capacity to deliver rich, comprehensible, and multimodal input that far exceeds what is typically accessible in conventional classroom settings. Whereas textbook-based materials often present decontextualized, sanitized, and artificially simplified language samples, AI-VEs embed input within visually realistic, context-saturated environments. This ecological validity supports comprehension by fusing linguistic forms with dense networks of extralinguistic cues, thereby strengthening form–meaning connections and reducing ambiguity.

Within AI-VEs, input is not confined to a single channel but is distributed across multiple, partially redundant semiotic modes, including spoken language from virtual agents, environmental visuals and spatial configurations, textual annotations and subtitles, gestures, gaze, facial expressions, and a variety of background sounds and other paralinguistic signals. This multimodal input enables learners to interpret meaning through interconnected sensory cues, thereby enhancing comprehension and reducing cognitive load.

From a cognitive perspective, such multimodal stimulation aligns with dual-coding and multimedia learning theories, which posit that information encoded through multiple channels is more robustly processed and retained. Learners can draw on contextual and visual affordances to infer meaning, thereby reducing intrinsic cognitive load and rendering input more comprehensible, even when its linguistic complexity is relatively high.

Beyond richness and realism, AI-VEs adjust the difficulty of input dynamically through natural language understanding and ongoing estimates of learner proficiency. When the system detects signs of comprehension difficulty—for instance, through response latency, error patterns, or explicit learner signals—it can: slow down speech rate, simplify syntactic structures, highlight relevant referents in the visual field, provide paraphrased or elaborated explanations.

This adaptive modulation keeps input within the learner's "i+1" zone, maintaining an optimal balance between challenge and comprehensibility. In doing so, AI-VEs transform Krashen's notion of comprehensible input from a largely teacher-driven construct into a finely calibrated, algorithmically managed process.

A small-scale pilot study was conducted with 12 Chinese EFL undergraduates who completed two AI-VE interaction tasks lasting 25 minutes each. Learners' oral production was transcribed and analyzed for lexical sophistication, including mid-/low-frequency type counts and lexical density. Results showed significant improvement between pre- and post-task performances, with the average number of mid-/low-frequency lexical items increasing from 3.1 to 6.4 types, and lexical density rising from 41% to 48%. Interview data revealed that multimodal cues—such as avatar gestures, environmental affordances, and onscreen lexical prompts—supported learners' word recognition and meaning-making. One participant noted, "I could guess new words more easily because the environment showed what they meant." These findings suggest that the multimodal and contextualized nature of AI-VE input can effectively enhance lexical depth and support comprehension, aligning with theories of comprehensible and multimodal input.

Table 1 Effects of Multimodal Input on Lexical Development

Indicator	Pre-task Mean (M)	Post-task Mean (M)	Change (Δ)	Description
Mid-/Low-frequency lexical types	3.1	6.4	+3.3	Increased use of less frequent lexical items
Lexical density (%)	41%	48%	+7%	Higher lexical richness in post-task performance
Learner reports on multimodal support	—	—	Mentioned by 8 learners	Learners perceived benefits of gestures, visuals, and spatial cues

5.2 Deepened interaction and negotiation of meaning

Interaction is widely recognized as central to language development, and AI-VEs create dense opportunities for learners to engage in real-time negotiation of meaning. Rather than remaining passive recipients of language, learners participate in reciprocal exchanges with virtual agents and, where appropriate, with peers embedded in the same immersive environment.

Advances in NLP and dialogue management allow AI-driven agents to interpret learner input—both spoken and written—and respond in ways that are contextually appropriate and interactionally contingent. Typical patterns include: clarification requests ("Do you mean…?"), comprehension checks ("Did you say…?"), recasts ("You mean the departure gate, not leaving door."), elaboration prompts ("Could you explain why?").

These moves closely parallel negotiation sequences documented in human–human interaction and serve a similar developmental function: they increase the salience of form–meaning mismatches and stimulate deeper processing, thereby facilitating interlanguage restructuring.

AI-VEs embed such interaction within task-based communicative episodes that require sustained engagement, such as booking a hotel room, participating in a group meeting or project, resolving a cross-cultural misunderstanding, explaining a process or providing advice.

These tasks compel learners to mobilize linguistic, pragmatic, and strategic resources over extended turns, increasing both the quantity and the qualitative depth of interaction. The resultant negotiation of meaning contributes not only to linguistic noticing but also to conceptual enrichment and pragmatic development.

Interaction logs were analyzed following Long's (1996) negotiation categories. Across 96 minutes of learner–avatar dialogue, the AI-VE system generated 128 negotiation moves, including clarification requests (37%), confirmation checks (29%), and recasts (21%). Learners demonstrated 82% successful uptake, producing self-repairs, reformulations, or extended responses. For example, when a learner said, "I took the wrong bus," the avatar responded, "Do you mean you got on the wrong bus?" prompting a modified output. The density and responsiveness of these interactional moves illustrate the AI-VE's ability to simulate negotiation-rich communicative environments, providing learners with repeated opportunities for noticing form–meaning mismatches—central to interaction-driven SLA.

Table 2 Interactional Negotiation Moves in AI-VE

Indicator	Count	Percentage	Description
Total interaction time	96 minutes	—	Combined learner–avatar interaction duration
Total negotiation moves	128	100%	All coded negotiation moves
Clarification requests	47	37%	Requests for clarification by avatar
Confirmation checks	37	29%	Avatar verification of learner intent
Recasts	27	21%	Corrective reformulations provided by avatar
Learner successful uptake	105	82%	Learner reformulation or repair following negotiation

5.3 Promotion of routinization and automatization

A further mechanism concerns the routinization of language use, whereby learners develop stable, easily retrievable linguistic patterns that support rapid, fluent communication. AI-VEs promote such routinization by orchestrating repeated exposure to recurring communicative functions and formulaic sequences across diverse situational contexts.

5.3.1 Repetition and variability

Learners repeatedly encounter similar communicative demands—such as making requests, giving directions, or expressing disagreement—across multiple scenarios. It is precisely this interplay of repetition with contextual variability that fosters robust routines. Through cyclical engagement with analogous tasks, learners gradually internalize: lexical chunks and formulaic expressions, frequent collocations, pragmatic formulas and discourse markers, idiomatic and semi-fixed expressions. These routines free up attentional resources, enabling learners to allocate more cognitive capacity to higher-order planning and meaning negotiation.

5.3.2 Immediate corrective feedback

Routinization is additionally reinforced by immediate, fine-grained corrective feedback generated by AI systems (Zheng & Yu, 2018). In contrast to delayed or global feedback typical of traditional instruction, AI-VEs can: highlight specific morphological or syntactic errors, suggest more natural or context-appropriate phrasings, propose pragmatically sensitive alternatives, offer frequency-based lexical recommendations. These tightly coupled feedback loops support automatization by preventing the consolidation of erroneous patterns and by repeatedly reinforcing more target-like forms at the moment of use (Lyster & Ranta, 1997).

5.3.3 Empirical illustration: routinization and formulaic sequence development

Analysis of pre- and post-task learner speech showed marked increases in the use of formulaic sequences. The mean frequency of formulaic expressions per 100 words rose from 7.8 to 13.2, with notable gains in service encounters (e.g., "I'd like to…," "Could you please…") and discourse markers (e.g., "actually," "the thing is…"). Participants commented that repeated exposure to similar communicative functions across varied VR scenarios led to "hearing the same phrases again and again but in different situations," enabling recognition, retention, and eventual automatization. These results support claims that repetition + contextual variability fosters routinization, contributing to increased fluency and formulaic competence.

Table 3 Changes in Formulaic Sequence Use

Indicator	Pre-task (per 100 words)	Post-task (per 100 words)	Change (Δ)	Description
Formulaic sequence frequency	7.8	13.2	+5.4	Increased use of formulaic expressions
Service-encounter expressions	2.1	4.4	+2.3	Greater use of situationally appropriate routines
Discourse markers	1.4	2.6	+1.2	More natural discourse flow

5.4 Personalized and adaptive learning pathways

Personalization constitutes one of the most distinctive affordances of AI-based environments. Rather than imposing a uniform sequence of learning activities, AI-VEs continuously monitor learner behavior and adapt instructional trajectories to individual needs.

AI systems construct evolving learner models by analyzing: response accuracy and error types, processing speed and hesitation patterns, vocabulary range and lexical sophistication, interactional moves (e.g., turn-taking, repair, initiative).

These data feed into adaptive algorithms that infer each learner's strengths, weaknesses, and preferred modes of engagement. The resultant profile becomes the basis for tailored instructional interventions.

On this basis, tasks are sequenced according to principles analogous to Vygotsky's Zone of Proximal Development (ZPD), with the system: gradually increasing linguistic and cognitive complexity, introducing new pragmatic and discourse-level challenges, varying communicative roles (e.g., from initiator to responder, novice to expert), scaffolding autonomy by progressively reducing support.

This adaptivity enhances the efficiency of learning and mitigates the risks of both cognitive overload and disengagement, allowing learners to remain in a state of optimally productive challenge.

Platform logs revealed that the AI-VE system adjusted task difficulty according to learner performance. For students who consistently met task goals, the avatar's speech rate increased from 120 wpm to 145 wpm, while lexical density increased by 18%. Conversely, when learners hesitated for more than 5 seconds, the system automatically provided simplified paraphrases or highlighted relevant visual referents. Learners perceived these adjustments positively, describing the experience as "not too hard, not too easy," and "like the system understands when I need help." These findings demonstrate how AI-VEs can generate dynamic, data-driven scaffolding aligned with Vygotsky's ZPD, supporting individualized and productive learning trajectories.

Table 4 Adaptive Adjustment Patterns in AI-VE

Indicator	Initial Level	Adjusted Level	Change (Δ)	Description
Avatar speech rate (wpm)	120 wpm	145 wpm	+25 wpm	Increased difficulty based on learner performance
Lexical density of system prompts	0.39	0.46	+18%	More complex input for advanced learners
Automatic simplification events (triggered by >5s hesitation)	19	—	—	Context-sensitive scaffolding initiated by system

5.5 Emotional engagement and motivational support

Affective variables are now widely recognized as critical mediators of language learning outcomes. AI-VEs, by virtue of their immersive, interactive, and often game-like nature, have substantial potential to positively shape learners' emotional and motivational states (Kang & Han, 2021).

Virtual environments create low-stakes communicative spaces in which learners can experiment with language without fear of social embarrassment or evaluative judgment (Fredrickson, 2001). This safety enables learners to: take communicative risks, attempt more complex structures, self-correct overtly, tolerate ambiguity and temporary breakdowns (Makransky & Lilleholt, 2018). Reduced anxiety is associated with higher willingness to communicate (WTC), increased output, and more frequent engagement in interactional work (MacIntyre et al., 1998).

AI-VEs can also enhance intrinsic motivation through narrative and game design, for example by:

embedding tasks within compelling storylines, enabling role-play with meaningful identities, providing reward systems and progress indicators, offering visually and aurally engaging worlds. Motivated learners are more likely to persist, to explore beyond minimal task requirements, and to engage in deeper cognitive processing, all of which are conducive to sustained development.

More advanced AI systems incorporate elements of affective computing, enabling virtual agents to detect emotional cues—such as frustration, confusion, or disengagement—and respond empathetically (Dewaele, 2013).

They may: adjust task difficulty or pace, provide encouragement or reassurance, offer additional scaffolding when needed. This socio-emotional responsiveness fosters a more supportive learning climate and strengthens learners' sense of presence, relatedness, and self-efficacy (Pekrun & Linnenbrink-Garcia, 2012).

A brief post-experience questionnaire measuring Willingness to Communicate (WTC) and Perceived Presence (α = .89) revealed substantial increases after the AI-VE tasks. Mean WTC scores rose from 3.2 to 4.4 (out of 5), and perceived presence reached 4.6, indicating strong immersion. Learners reported reduced anxiety due to the "safe, non-judgmental environment," with one student explaining, "I'm not afraid to make mistakes because the avatar doesn't judge me." These results highlight the affective benefits of AI-VEs, showing strong potential for enhancing motivation, risk-taking, and communicative willingness—essential affective variables in SLA.

Table 5 Learners' Affective and Perceptual Outcomes in the AI-Based Virtual Environment (AI-VE)

Variable	Pre-task Mean (M)	Post-task Mean (M)	Change (Δ)	Reliability (α)	Description
Willingness to Communicate (WTC)	3.2	4.4	+1.2	.89	Increased willingness to speak in AI-VE
Perceived Presence	—	4.6	—	.86	Strong sense of immersion
Reported anxiety reduction	—	—	Mentioned by 9 learners	Learners felt safer and less judged	

5.6 Development of intercultural competence

In an increasingly globalized communicative landscape, intercultural competence is a crucial dimension of communicative competence. AI-VEs are particularly well suited to orchestrate intercultural encounters that would be logistically and ethically difficult to stage in traditional classrooms.

Virtual scenarios can be designed to reflect culturally specific norms, practices, and expectations. Learners may observe and enact: culturally appropriate greetings and leave-takings, politeness strategies and facework routines, gesture and proxemics conventions, culturally contingent conversational styles and turn-taking norms.

Through such experiences, learners gradually develop pragmatic sensitivity and cultural literacy.

AI-VEs also provide spaces in which learners can safely encounter and work through intercultural misunderstandings. Scenarios might involve: misaligned expectations about directness or politeness, divergent interpretations of the same behavior, contrasting conflict resolution styles.

Feedback and guided reflection help learners understand the cultural logics at play and adjust their communicative behavior accordingly, fostering critical intercultural awareness.

Over time, repeated engagement with diverse interlocutors, perspectives, and value systems contributes to broader global competence. Learners develop: empathy and openness to difference, tolerance for ambiguity, interpretive skills in cross-cultural contexts, readiness to participate in international academic and professional communication.

In this sense, AI-VEs function not only as language learning tools, but also as environments for cultivating dispositions and skills associated with global citizenship.

In sum, AI-based virtual environments promote language development through a tightly interwoven set of cognitive, interactional, affective, and sociocultural mechanisms. They enrich and individualize input, intensify opportunities for interaction and negotiation of meaning, accelerate routinization and automatization, support adaptive learning trajectories, sustain motivation and emotional engagement, and scaffold the emergence of intercultural competence. Collectively, these mechanisms furnish the empirical and theoretical grounding for the three-layer language development model elaborated in the subsequent section.

Learners' scores on a short version of the Intercultural Sensitivity Scale (ISS) increased moderately in the subscales of interaction engagement (Δ = +0.38) and respect for cultural difference (Δ = +0.41). Interview data indicated enhanced awareness of pragmatic and cultural variation. For example, a student reflected, "Now I understand why English speakers prefer more direct expressions—it's not rude; it's their cultural style." These findings demonstrate the potential of AI-VEs to serve as safe intercultural rehearsal spaces, supporting pragmatic awareness and intercultural communicative competence.

Table 6 Development of Intercultural Competence

ISS Subscale	Pre-task Mean (M)	Post-task Mean (M)	Change (Δ)	Description
Interaction Engagement	3.41	3.79	+0.38	Greater willingness to engage interculturally
Respect for Cultural Difference	3.52	3.93	+0.41	Enhanced cultural awareness and sensitivity
Pragmatic Awareness (Interview-based)	—	—	Reported by 7 learners	Increased awareness of pragmatic differences

6. The Three-Layer AI-VE Learning Model

Building on the theoretical foundations and mechanisms elaborated in the preceding sections, this study advances a three-layer developmental model that explicates how AI-based virtual environments (AI-VEs) can support Chinese EFL learners' linguistic, communicative, and intercultural growth in a systematic and theoretically principled manner. The model comprises three analytically distinct yet dynamically interrelated layers: (1) an outer layer representing the ecosystem of immersive virtual scenarios, (2) a middle layer representing the Input–Interaction–Output–Reflection (IIOR) learning cycle, and (3) an inner layer representing the progressive development of learner competencies. These layers do not operate in isolation; rather, they form a coupled system in which contextual affordances, cognitive–interactional mechanisms, and developmental processes mutually shape one another, thereby sustaining complex, nonlinear trajectories of language growth.

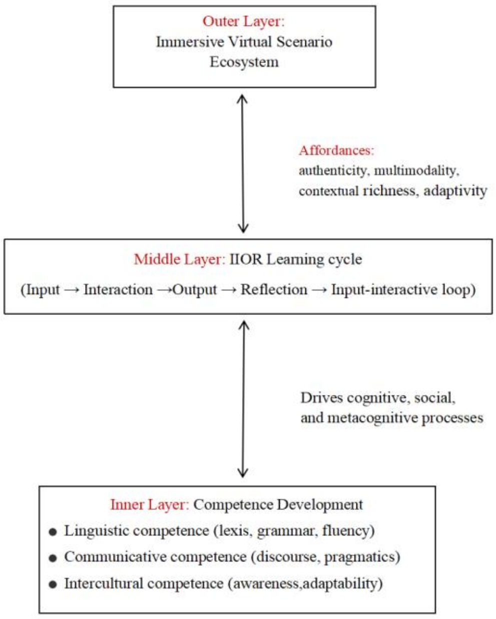

Figure 2 Three-Layer Model

6.1 Immersive virtual environment ecosystem

The outer layer captures the ecology of immersive, contextually rich scenarios that AI-VEs make available. It can be understood as the "situated world" in which learning is enacted—a world that anchors linguistic forms, pragmatic norms, and communicative goals in realistic, experientially meaningful contexts.

6.1.1 Diversity of communicative scenarios

AI-VEs can orchestrate a broad spectrum of scenarios spanning everyday, academic, professional, and cross-cultural domains, such as: Daily life scenarios: shopping, dining, transportation, healthcare encounters; Academic scenarios: group discussions, seminars, oral presentations, office-hour consultations; Professional scenarios: job interviews, team meetings, negotiations, collaborative problem-solving; Cross-cultural scenarios: interaction with culturally diverse avatars, intercultural misunderstandings, conflict mediation.

By exposing learners to this diversity of settings, AI-VEs familiarize them with varied pragmatic norms, conversational styles, and linguistic registers, and accommodate learners with different proficiency levels and learning goals. The breadth of scenarios serves as a repertoire of "practice worlds" in which learners can rehearse, refine, and transfer communicative skills.

6.1.2 Authenticity and ecological validity

Authenticity is central to the pedagogical value of the outer layer. Physical realism—manifested in 3D environments (Dalgarno & Lee, 2010), embodied avatars, and naturalistic soundscapes—interacts with sociolinguistic realism—reflected in culturally appropriate speech acts, social hierarchies, and politeness conventions—to create ecologically valid communicative situations. Such authenticity narrows the frequently cited "classroom-to-world gap" by approximating the perceptual, social, and pragmatic conditions under which language is used outside the classroom. In doing so, it enhances the likelihood that skills acquired within AI-VEs will transfer to real-world communicative contexts (Bailenson, 2018).

6.1.3 Affordances unique to AI-driven virtual environments

Beyond authenticity, AI-VEs afford several pedagogical possibilities that go beyond both conventional classrooms and non-AI VR systems: Unlimited repetition and practice opportunities, unconstrained by timetable or teacher availability; Consistent access to interlocutors, alleviating human resource limitations; Real-time intelligent feedback, integrated seamlessly into ongoing interaction; Dynamic adaptivity, through which scenarios respond to learner behaviors and proficiency; Psychological safety, allowing learners to take communicative risks without social penalties.

In this sense, the outer layer furnishes the contextual platform upon which more fine-grained cognitive, interactional, and affective processes can operate.

6.2 Input–Interaction–Output–Reflection learning cycle

At the core of the model lies the middle layer, which represents the IIOR learning cycle that mediates between environmental affordances and internal developmental processes. This layer operationalizes SLA principles by structuring learning as a recursive sequence of input, interaction, output, and reflection, each stage amplifying the others.

As learners navigate virtual scenarios, they encounter multimodal, context-embedded input. This input: carries authentic linguistic and pragmatic cues, is supported by environmental affordances that facilitate meaning inference, is more memorable due to its sensory and situational richness, can be processed through different channels to accommodate diverse cognitive preferences.

Because such input is responsive to learner performance—becoming more complex, more scaffolded, or more elaborated as needed—it supports gradual, adaptive scaffolding rather than static presentation.

Interaction constitutes the second stage of the cycle. AI-driven avatars can converse with learners, pose challenges, and probe their understanding in ways that stimulate elaborated discourse. Through comprehension checks, clarification requests, negotiation moves, and discourse-level engagement, learners are pushed to process input more deeply, test their hypotheses about form–meaning mappings, and refine their interlanguage. Interaction simultaneously exposes them to sociolinguistic and pragmatic norms, enabling them to align their language use with context-sensitive expectations.

Output is not merely an end-product but an integral developmental process. In AI-VEs, learners must articulate meanings using linguistic and strategic resources across multiple modalities, including: spoken utterances, written messages, paralinguistic actions (e.g., gesture selection, response choices), multimodal explanations combining verbal and visual means.

Such performance conditions encourage learners to formulate, monitor, and revise their language, thereby testing emerging hypotheses and restructuring underlying linguistic knowledge.

Reflection differentiates AI-VEs from many other learning environments. AI systems can generate instantaneous, fine-grained, and personalized feedback, enabling learners to engage in: noticing and heightened awareness of linguistic and pragmatic features, targeted error correction, metacognitive evaluation of strategies and performance, strategic adjustment of learning approaches.

Through dashboards, visualizations of performance trajectories, and replayable interaction records, learners gain tools for self-monitoring and self-regulation, which are essential for long-term retention and autonomous learning.

The IIOR cycle is recursive and dynamic rather than linear. Learners repeatedly move through input, interaction, output, and reflection, with each iteration deepening cognitive engagement and contributing to cumulative development. This dynamic, non-linear progression is consistent with Dynamic Systems Theory perspectives on SLA, which emphasize variability, emergence, and sensitivity to contextual conditions. The middle layer thus captures the temporal, processual dimension of learning within AI-VEs.

6.3 Progressive development of language competence

The innermost layer represents the developmental outcomes that emerge from the interaction of environmental affordances and the IIOR cycle. Here, language growth is conceptualized as a staged, cumulative progression across three interrelated domains: lexicalization, communicative competence, and intercultural competence.

6.3.1 Lexicalization and formulaic language development

The first stage involves the consolidation of lexical chunks, fixed expressions, and formulaic sequences. Because fluent communication disproportionately relies on pre-fabricated language rather than online rule construction, this stage is foundational. AI-VEs foster lexicalization by:

Repeatedly exposing learners to high-frequency formulas (e.g., requests, apologies, hedges), embedding these formulas in varied yet functionally similar contexts, reinforcing stable patterns through targeted feedback.

Over time, learners can retrieve these expressions with increasing speed and deploy them flexibly in communication, thereby freeing cognitive resources for higher-level processing.

6.3.2 Development of communicative competence

The second stage extends beyond lexical and grammatical knowledge toward fuller communicative competence, encompassing: Linguistic competence (vocabulary, grammar, phonology), Pragmatic competence (speech act realization, politeness strategies, appropriateness), Discourse competence (cohesion, coherence, turn-taking and topic management), Strategic competence (repair, paraphrase, circumlocution, managing breakdowns).

AI-VEs support the emergence of these competencies by placing learners in authentic interactional contexts where such skills are not abstract objectives but practical necessities for task completion and social alignment.

6.3.3 Development of intercultural competence

The highest stage is characterized by the development of intercultural communicative competence, increasingly indispensable in globalized communicative arenas. AI-VEs cultivate this competence by: exposing learners to diverse cultural behaviors, beliefs, and value orientations, simulating culturally sensitive or high-stakes interactions, providing feedback on culturally inappropriate or ambiguous behaviors, prompting critical reflection on learners' own assumptions and interpretive frames. Through such experiences, learners develop: cultural awareness and critical cultural consciousness, empathy and openness to difference, the ability to adapt communicative repertoires to shifting cultural expectations, competence in navigating and resolving intercultural misunderstandings.

6.4 Interrelationships among the three layers

The three-layer model is designed to be integrative rather than strictly hierarchical. The immersive environment at the outer layer shapes the opportunities and constraints under which the IIOR learning cycle unfolds. The learning cycle at the middle layer, in turn, drives the consolidation and transformation of competencies at the inner layer. As learners' competencies evolve, they perceive and exploit environmental affordances differently, engage in more complex and nuanced interactional work, assume more sophisticated roles within the virtual ecology. Thus, the relationship among the three layers is fundamentally dialogic: the outer layer provides the contextual affordances that shape learning opportunities, the middle layer organizes the cognitive, interactional, and reflective processes that drive the learning cycle, and the inner layer represents the emergent developmental outcomes that arise from learners' engagement with these processes. Together, these layers interact dynamically to support complex and continuous language development. Their continuous interplay supports transferability (from virtual to real contexts), scalability (across cohorts and settings), and sustainability (across time and proficiency levels).

The proposed three-layer model offers a comprehensive and theoretically grounded account of how AI-based virtual environments can foster language development among Chinese EFL learners. By integrating immersive contextual affordances (outer layer), a recursive IIOR learning mechanism (middle layer), and a staged progression of language competencies (inner layer), the model bridges SLA theory, educational technology design, and pedagogical practice. It underscores the potential of AI-VEs not only to address persistent structural challenges in Chinese EFL education—such as limited authentic input, constrained interaction, and insufficient pragmatic experience—but also to promote higher-order outcomes, including communicative effectiveness and intercultural sensitivity, that are crucial for participation in global communicative communities.

7. Educational and Social Implications

The three-layer model advanced in this study carries significant implications for EFL pedagogy, curriculum development, teacher professionalization, educational technology design, and broader issues of equity and global citizenship. Beyond its contributions to linguistic proficiency, the model foregrounds the development of intercultural competence, learner autonomy, and equitable access—outcomes essential in an increasingly digital and interconnected world. This section explicates how AI-based virtual environments (AI-VEs) can transform Chinese EFL education and facilitate sustainable language learning ecosystems.

7.1 Pedagogical implications

AI-VEs offer transformative potential for reconfiguring the pedagogical landscape of English education in China. The prevailing teacher-centered, textbook-driven paradigm often restricts communicative practice, fails to supply timely individualized feedback, and provides limited exposure to authentic language use. AI-VEs directly challenge these systemic constraints.

Immersive virtual scenarios reposition teachers from information transmitters to designers and facilitators of experiential learning. Rather than rehearsing contrived drills, learners engage in ecologically valid communicative episodes, such as: ordering food in a restaurant simulation, participating in academic debates or group discussions, navigating healthcare appointments, resolving intercultural misunderstandings. These scenarios align closely with Task-Based Language Teaching (TBLT) principles, embedding communication within meaningful, goal-oriented activity.

Given the prevalence of large, heterogeneous classrooms in China, differentiated instruction remains a persistent challenge. AI-VEs mitigate this issue through dynamic adaptivity, enabling learners to progress at individualized paces while teachers utilize system-generated analytics to deliver targeted instructional support.

AI-generated feedback is immediate, consistent, fine-grained, and available on demand. Automated dashboards visualize learners' development across lexical diversity, syntactic complexity, fluency, and interactional patterns. These affordances cultivate metacognitive awareness, encourage autonomous learning, and enable learners to assume a proactive role in monitoring their progress.

7.2 Curriculum and assessment implications

Integrating AI-VEs into curricular frameworks requires a fundamental reconceptualization of what constitutes meaningful learning content and how learning should be assessed.

Conventional curricula fragment language skills into discrete units; however, AI-VEs enable scenario-based learning, where skills are integrated holistically. Curricular redesign may involve: thematic modules (e.g., travel, academic communication), sociocultural modules (e.g., conflict resolution across cultures), task sequences that scaffold increasing communicative complexity. This approach aligns with pragmatic, interactional, and discourse-based models of competence.

AI-VEs generate large-scale, fine-grained performance data, making them ideal platforms for formative, performance-based assessment. Analytics allow educators to evaluate: fluency measures (pause distribution, speech rate), syntactic complexity indices, vocabulary diversity metrics, pragmatic appropriateness and politeness strategies, interactional competence indicators.

AI-generated portfolios could ultimately supplement or transform existing high-stakes exam-based assessment regimes.

7.3 Teacher professional development

The pedagogical integration of AI-VEs requires teachers to expand their technical and pedagogical repertoires. Professional development initiatives should enable teachers to: understand and navigate immersive technologies, interpret AI-generated learner analytics, design scenario-based communicative tasks, scaffold learner engagement within virtual interactions, manage hybrid AI–human instructional environments. Teachers equipped with Technological Pedagogical Content Knowledge (TPACK) will be well positioned to orchestrate productive AI-assisted learning experiences.

7.4 Technology and system design implications

From a design standpoint, the three-layer model provides a conceptual blueprint for constructing effective AI-VE systems: Outer Layer: Technology developers must create diverse, culturally grounded, pedagogically aligned scenarios.

Middle Layer: Systems should support the full Input–Interaction–Output–Reflection (IIOR) cycle, enabling adaptive dialogues, real-time feedback, and learning analytics.

Inner Layer: Platforms should track developmental trajectories, not merely isolated performance events (Shneiderman, 2020).

This integrated design philosophy ensures that AI-VEs function not merely as linguistic tools but as full-fledged learning ecosystems.

7.5 Equity and access

A key social contribution of AI-VEs lies in their capacity to democratize access to high-quality learning resources.

Learners in rural or under-resourced regions often lack access to proficient English teachers or authentic communicative contexts (Selwyn, 2019). AI-VEs can provide uniformly high-quality, immersive learning opportunities, regardless of geographic location or school resources (Floridi, 2019).

Because AI-driven interactions are nonjudgmental, flexible, and infinitely repeatable, they are particularly beneficial for learners experiencing: communication anxiety, low self-confidence, special educational needs. Thus, AI-VEs support inclusive education and broaden participation in EFL learning (Williamson & Eynon, 2020).

7.6 Intercultural competence and global citizenship

AI-VEs expose learners to diverse cultural norms, values, and communicative styles, allowing them to cultivate global competence. Through repeated participation in simulated intercultural encounters, learners develop: cultural empathy, pragmatic sensitivity, strategies for managing intercultural misunderstandings, readiness for international academic or professional engagement.

In this way, AI-VEs contribute to the formation of globally competent citizens equipped for cross-cultural interaction.

8. Conclusion

This study advances a comprehensive and empirically supported model illustrating how AI-based virtual environments (AI-VEs) can facilitate the language development of Chinese EFL learners. By integrating principles from Second Language Acquisition (SLA), constructivism, and situated learning, the proposed three-layer model conceptualizes language learning as a dynamic, emergent process shaped by immersive contextual affordances, iterative cognitive–interactional cycles, and progressive developmental trajectories. Importantly, the pilot evidence incorporated across six core mechanisms—multimodal input, interaction, routinization, adaptivity, affective support, and intercultural learning—provides initial empirical validation for the theoretical claims of the model.

Findings from the pilot implementation indicate that AI-VEs can meaningfully enhance learners' lexical sophistication, increase opportunities for negotiation of meaning, foster the use of formulaic language, generate personalized learning pathways, boost communicative confidence, and promote intercultural awareness. These results not only substantiate the conceptual model but also demonstrate how the Input–Interaction–Output–Reflection (IIOR) cycle can be operationalized through AI-driven, immersive learning environments. Together, the theoretical integration and empirical illustrations underscore the potential of AI-VEs to serve as powerful mediational tools for supporting multidimensional EFL development in the Chinese context.

8.1 Contributions of the study

This study makes several interrelated theoretical, empirical, pedagogical, and technological contributions.

Theoretically, it offers one of the first integrated frameworks that systematically connects established SLA mechanisms with the unique affordances of AI-driven virtual environments. The inclusion of empirical illustrations strengthens the explanatory power of the model and demonstrates how the IIOR cycle manifests in actual learner behavior.

Empirically, although exploratory in nature, the pilot findings provide initial support for key mechanisms within the model—lexical enrichment, negotiation of meaning, routinization, adaptivity, affective engagement, and intercultural development—thereby laying a foundation for future large-scale validation.

Pedagogically, the study provides actionable insights for teachers and curriculum designers seeking to integrate AI-VEs into communicative EFL instruction. The findings highlight the value of immersive, adaptive, and low-anxiety environments for promoting output, risk-taking, and metacognitive reflection.

Technologically, the model offers a design roadmap for developers by specifying how AI-driven features—such as adaptive feedback, multimodal input delivery, learning analytics, and scenario-based interaction—can be aligned with SLA principles.

Socially, the study underscores the potential of AI-VEs to democratize access to high-quality linguistic input and intercultural experiences, thereby supporting educational equity and global citizenship in China's evolving digital learning landscape.

8.2 Limitations of the study

The study is not without limitations. The empirical component constitutes a small-scale pilot involving a limited number of participants, which restricts the generalizability of the findings. The AI-VE platform examined in the pilot is only one instantiation of the broader ecological space of immersive learning technologies, and results may vary across platforms with different capabilities. Furthermore, the short duration of learner engagement does not allow for conclusions regarding long-term developmental trajectories or sustained proficiency gains.

In addition, the study primarily examines adult learners with intermediate proficiency; future research should explore diverse learner profiles, including younger learners, varying proficiency levels, and individuals with special educational needs. Finally, ethical considerations related to data privacy, learner profiling, and algorithmic adaptivity warrant closer examination as AI-mediated learning continues to expand.

8.3 Directions for future research

Building on the conceptual and pilot findings, several promising directions for future research are identified:

Large-scale empirical validation of the three-layer model, using experimental, quasi-experimental, or mixed-methods designs across varied educational settings.

Longitudinal analyses tracing changes in CAF measures, interaction patterns, and intercultural competence over extended periods.

Development of assessment instruments specifically tailored to AI-VE contexts, particularly for measuring intercultural competence, affective engagement, and real-time interactional competence.

Multimodal analytics, including eye-tracking, EEG, sentiment analysis, and behavioral telemetry, to model cognitive and affective processes during immersive language learning.

Cross-platform and cross-linguistic comparisons to determine the transferability of the model across technologies, languages, and cultural contexts.

Teacher–AI collaboration models, exploring how educators can effectively orchestrate hybrid classrooms that integrate human pedagogy with AI-driven scaffolding.

Collectively, these research avenues will help refine the model, strengthen its empirical grounding, and expand its applicability across diverse EFL learning environments.

Conflict of Interest

The authors declare that they have no conflicts of interest to this work.

Acknowledgement

This research was funded by: Y202557459 (A Study on the Model of Promoting Language Proficiency Development Among Chinese EFL Learners Through AI-Driven Virtual Scenarios).

References

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. W. W. Norton & Company.

Chapelle, C. A. (2009). The relationship between second language acquisition theory and computer-assisted language learning. *The Modern Language Journal, 93*(s1),741-753.

Chen, X., & Teo, T. (2021). Effects of virtual reality on EFL speaking performance: A meta-analysis. *Interactive Learning Environments*. Advance online publication.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7),13-20.

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1),10-32.

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp.94-112). Routledge.

Dewaele, J.-M. (2013). Emotions in multiple languages. *Studies in Second Language Learning and Teaching, 3*(2), 297-324.

Dörnyei, Z. (2005). *The Psychology of the Language learner: Individual differences in second language acquisition*. Lawrence Erlbaum Associates.

Engeström, Y. (1987). *Learning by expanding: An* *activity-theoretical approach to developmental research*. Orienta-Konsultit.

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*(3), 218-226.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Lawrence Erlbaum Associates.

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI and code. *Language Learning & Technology, 27*(1),4-26.

Graesser, A. C. (2016). Conversations with Auto Tutor help students learn. *International Journal of Artificial Intelligence in Education, 26*(1),124-132.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Housen, A., Kuiken, F., & Vedder, I. (Eds.).(2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins Publishing Company.

Hu, Z., & McGrath, I. (2011). Integrating ICT into college English teaching in China: A case study. *Computer Assisted Language Learning, 24*(5),1-20.

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies,23*(4), 1515-1529.

Jiang, L., & Wang, Y. (2023). Large language models in second language acquisition: A critical review. *Language Teaching Research*. Advance online publication.

Kang, M., & Han, S. (2021). A meta-analysis of the effects of affective factors in intelligent tutoring systems. *Computers & Education, 173*, 104288.

Lan, Y. J. (2020). Immersion, interaction, and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning & Technology, 24*(1),1-15.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590–619.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Li, J., & Ni, X. (2023). Exploring the use of ChatGPT in EFL writing: A case study of Chinese university students. *Computer Assisted Language Learning*. Advance online publication.

Li, Z., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1-18.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp.413-468). Academic Press.

Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour, 1*,0028.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*(1),37-66.

MacIntyre, P D., Dörnyei, Z., Clément, R., & Noels, K. A.(1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal, 82*(4),545-562.

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional engagement in virtual reality. *Computers in Human Behavior, 87*,225-233.

Parmaxi, A., & Zaphiris, P. (2017). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments, 25*(8),1-13.

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement.In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp.259-282). Springer.

Pellas, N. (2014). The influence of computer self-efficacy, metacognitive self-regulation and self-esteem on student engagement in online learning programs: Evidence from the virtual world of Second Life. *Computers in Human Behavior, 35*,157-170.

Peterson, M. (2012). EFL learner collaborative interaction in Second Life. *ReCALL, 24(*1),20-39.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, L. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2),582-599.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction, 36*(6), 495-504.

Swain, M. (1995). Three functions of output in the second language classroom. In G.Cook & B. Seidlhofer (Eds.), Principle and practice in applied linguistics: *Studies in honour of H.G. Widdowson* (pp.125-144). Oxford University Press.

VanPatten, B. (2015). Input processing in adult SLA. In B. VanPatten & J.Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp.113-134). Routledge.

Vygotsky, L. S. (1978). *Mind in Society: The development of higher psychological processes*. Harvard University Press.

Wang, Y., & Heffernan, N. T. (2012). The “Assistance Model”: Leveraging how many questions and how much hint. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems* (pp.365-371). Springer.

Wang, Y., & Vásquez, C. (2014). The effect of target language use in social media on intermediate-level Chinese EFL learners' writing performance. *CALICO Journal, 31*(1),1-20.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Cambridge University Press.

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223-235.

Yeh, H.C., & Lan, Y. J.(2018). Fostering student teachers' EFL pragmatic awareness and performance through virtual reality. *Educational Technology & Society, 21*(2), 104-117.

Zhang, S. (2020). A systematic review of automated writing evaluation in ESL/EFL writing. *Journal of Computer Assisted Learning, 36*(5),1-15.

Zhao, Y. (2005). The future of research in technology and second language education. In Y. Zhao (Ed.), *Research in technology and second language learning: Developments and directions* (pp.445-457). Information Age Publishing.

Zheng, B., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese university students. *Assessing Writing,* 37,13-24.

How to Cite: Zou, J. & Ma, B. (2025). AI-Based Virtual Environments for Chinese EFL Learners: A Three-Layer Developmental Model with Pilot Evidence. *Contemporary Education and Teaching Research*, *06*(12), 532-554. https://doi.org/10.61360/BoniCETR252019451201