Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 5 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i05.60729

PAPER

Advancing Interactive English Learning through Mixed Reality

Jingjie Zhang[1], **Youzi Liang**[2] (✉)

[1]Hebei University of Water Resources and Electric Engineering, Cangzhou, China

[2]Hebei Vocational University of Industry and Technology, Shijiazhuang, China

liangyouzi819@163.com

ABSTRACT

Traditional English instruction has long been constrained by limited interactional forms, the absence of authentic learning contexts, and low efficiency in skill transfer, thereby hindering second language acquisition (SLA). Mixed reality (MR), with its capacity for multimodal interaction, offers a potential solution to these limitations. This study investigated the scenario adaptability and skill specificity of MR-based interaction, drawing on SLA theory and multimodal interaction theory. A randomized controlled trial with a mixed-methods design was conducted with 90 non-English-major university students over a 16-week intervention. One control group received traditional multimedia instruction, while three experimental groups engaged in MR-based interaction designs: (a) virtual characters with multimodality, (b) spatial scenarios with peer collaboration, and (c) personalized, multimodal interaction. Data were collected through language proficiency assessments, interaction behavior coding, immersion scales, and semi-structured interviews and were analyzed using repeated-measures analysis of variance (ANOVA) and thematic coding. Results showed that MR-based interaction significantly improved English proficiency, learning motivation, and perceived immersion compared with traditional instruction, with effects strengthening over time. Distinct patterns of interaction–skill adaptability emerged: virtual character interaction yielded the greatest gains in speaking, spatial scenarios most enhanced listening, and personalized adaptation produced the strongest improvements in writing. Immersion and depth of meaning negotiation fully mediated the relationship between MR interaction and learning outcomes. Learners with lower initial proficiency benefited more, and designs with lower cognitive load further amplified learning effects. A three-dimensional interaction format–scenario–skill adaptability model was developed, offering a mechanistic account of how multimodal interaction operates within SLA and providing a precision-oriented design paradigm for technology-enhanced language education. The findings also deliver theoretical and practical foundations for the scalable deployment of MR in educational settings.

KEYWORDS

mixed reality (MR), second language acquisition (SLA), multimodal interaction, randomized controlled trial (RCT), adaptability model, English education

Zhang, J., Liang, Y. (2026). Advancing Interactive English Learning through Mixed Reality. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(5), pp. 55–69. https://doi.org/10.3991/ijim.v20i05.60729

Article submitted 2025-08-30. Revision uploaded 2025-12-22. Final acceptance 2026-01-05.

1 INTRODUCTION

In an increasingly globalized environment, English proficiency has been regarded as a core competency for intercultural collaboration [1, 2]. However, the quality of English instruction has long been constrained by traditional pedagogical models, resulting in three structural limitations. First, interaction has predominantly relied on teacher–student questioning and text-based interpretation, leaving multisensory engagement severely lacking [3, 4]. Second, instructional contexts have remained disconnected from authentic language use situations, thereby hindering learners' ability to transfer linguistic knowledge to real-world communication [5, 6]. Third, skill training has typically followed a uniform instructional pattern, with limited alignment between interactional formats and the differentiated requirements of various skills [7, 8]. These constraints have collectively reduced the efficiency of SLA and have limited the ability of instructional models to meet contemporary expectations for comprehensive English competence. MR [9–11], which integrates physical environments with virtual content, offers a promising avenue for addressing these limitations. Through capabilities such as gesture tracking, speech interaction, and spatial immersion, MR enables an integrated learning ecology characterized by multimodal input, real-time interaction, and contextual transfer, providing an innovative solution for overcoming the deficits. Research on MR within language education has expanded notably in recent years [12, 13], and its core value has gained growing scholarly attention.

A review of leading publications from the past five years reveals that the application of MR in language education has evolved from early-stage vocabulary recall and grammar visualization toward more complex scenarios, including situated dialogue and intercultural communication. MR has been found to significantly enhance learning motivation [14], and its role in strengthening contextual transfer has been increasingly substantiated [15]. Nonetheless, three notable limitations continue to restrict the realization of MR's full instructional potential. First, earlier studies were limited by high equipment costs and small sample sizes, and research was predominantly led by computer science scholars, resulting in insufficient integration of educational and linguistic perspectives. Consequently, interaction design was often misaligned with pedagogical objectives, producing a fragmented research landscape. Second, research on adaptability remains underdeveloped. Most studies have evaluated isolated interactional formats—most commonly virtual character dialogue—without investigating the alignment between different MR interaction types and specific language skills (e.g., listening, speaking, reading, writing, and translation) or establishing criteria for selecting optimal interactional forms across instructional contexts. Third, evaluation systems have remained narrow in scope, relying primarily on language test scores while neglecting systematic assessment of interactional-process quality, such as the depth of meaning negotiation, as well as psychological constructs, including immersion and anxiety. This has limited the field's ability to identify the mechanisms by which MR exerts its pedagogical effects. Grounded in multimodal interaction theory and classroom interaction taxonomies, the core dimensions of interactive experience in English education can be characterized as interaction channels, interaction objects, and interaction depth. Multimodal interaction and deep meaning negotiation have been demonstrated to substantially enhance SLA efficiency [16]. However, empirical evidence explaining how MR optimizes these dimensions to improve learning outcomes remains insufficient.

To address the research gaps identified above, the present study adopted adaptability as its central logic and systematically compared the effects of different MR interactional formats across diverse skill-oriented instructional scenarios. Through this approach, an interaction format–scenario–skill adaptability model was constructed to remedy the lack of systematicity and mechanistic explanation in existing research. Specifically, three core questions were examined: (a) whether distinct MR interactional formats—such as virtual character collaboration and spatial scenario immersion—produce differential effects on various English language skills; (b) what the adaptability boundaries between interactional formats and instructional scenarios are, that is, which MR interactional format yields the greatest learning effect within a given scenario; and (c) whether experiential variables such as immersion and the depth of meaning negotiation mediate the relationship between MR interactional formats and learning outcomes and whether cognitive load and learners' initial language proficiency moderate these mediating processes.

The structure of the study follows a clear analytical logic. An integrated theoretical framework and research model are first established. This is followed by a detailed account of the research design, including participant selection procedures, intervention design, measurement instrument selection, and data analysis methods. The empirical results are subsequently presented, including the main effects, interaction effects, and mediation and moderation effects. These findings are then interpreted in light of existing research to articulate their theoretical contributions and practical implications. Finally, core conclusions are summarized, limitations are discussed, and directions for future research are outlined, providing a comprehensive scholarly foundation for the systematic application of MR in language education.

2 THEORETICAL FOUNDATIONS AND RESEARCH FRAMEWORK

The theoretical framework constructed in this study is not a simple accumulation of independent theories but an integrated, multilayered structure in which each theoretical component fulfills a clearly defined and complementary function within the MR–based English instructional process. Multimodal interaction theory serves as the foundational design rationale. Its central proposition—that human information processing occurs through multiple channels such as visual, auditory, and kinesthetic modalities—aligns closely with the multisensory cognitive requirements of language learning. This alignment provides the core logic for the design of MR interactional formats involving gesture tracking, speech interaction, and spatial movement. Situated learning theory constitutes the basis for environmental construction. The theory posits that learning must be embedded within authentic task-based contexts. The spatial scenario immersion afforded by MR enables the creation of "learning-in-action" language use environments, effectively addressing the decontextualization that characterizes traditional instruction and offering strong theoretical support for instructional scenario design. SLA theory provides the cognitive mechanism underpinning the framework. Krashen's input hypothesis emphasizes comprehensible input delivered within low-anxiety environments; MR-based interaction has been shown to reduce learners' speaking anxiety, thereby enhancing input absorption. Long's interaction hypothesis identifies meaning negotiation as a critical driver of language acquisition. MR-based virtual character collaboration and peer collaboration interactions furnish abundant opportunities for meaning negotiation, clarifying the core language internalization processes. The technology acceptance model (TAM) and cognitive load theory function jointly as the basis for moderating factors. Perceived usefulness and perceived ease of use, as delineated in TAM, together with the cognitive load constructs articulated in Sweller's framework, provide a systematic explanation for individual differences in learners' acceptance and adaptation to MR technology, thereby establishing a robust theoretical foundation for the subsequent analysis of moderation effects.

Based on the integrated theoretical framework, a research model incorporating an independent variable, a mediating variable, a dependent variable, and moderating variables was constructed. Logical linkages among these variables form a complete closed loop. The independent variable is the MR interactional format, operationalized through three composite schemes: (a) virtual character collaboration combined with multimodal interaction, (b) spatial scenario immersion combined with peer collaboration, and (c) personalized adaptation combined with multimodal interaction. The mediating variable is an interaction experience, comprising three core dimensions: engagement, immersion, and the depth of meaning negotiation. The dependent variable is English learning performance, assessed through standardized measures of five language skills—listening, speaking, reading, writing, and translation. The moderating variables include learners' initial language proficiency and cognitive load. Language proficiency is categorized into high, medium, and low levels, whereas cognitive load arises from the complexity of interactional formats and is classified into high and low levels. Grounded in the logical relationships among these variables, three core hypotheses were proposed in this study:

H1: MR interactional formats exert a significant positive effect on English learning performance, and the magnitude of this effect differs across distinct MR interactional formats.

H2: Interaction experience mediates the relationship between MR interactional formats and English learning performance.

H3: Learners' language proficiency and cognitive load moderate the effects of MR interaction; greater improvements are expected for learners with lower initial proficiency and within low–cognitive-load conditions.

3 RESEARCH METHODOLOGY

3.1 Research design

A mixed-methods design combining randomized controlled trials and subgroup comparisons was employed to ensure the scientific rigor and explanatory strength of the findings. A stratified random sampling procedure was used to assign 90 participants into four groups: one traditional multimedia control group and three MR interaction experimental groups, each consisting of 30 learners. Stratification was based on College English Test Band 4 scores and learning motivation levels, ensuring equivalent baselines in language proficiency and learning initiative. ANOVA confirmed that no statistically significant differences existed among the groups prior to the intervention, satisfying the homogeneity requirements of randomized controlled trials.

3.2 Participants

A total of 90 second-year university students majoring in fields other than English were selected as participants. To avoid ceiling and floor effects that might compromise validity, three inclusion criteria were established: (a) College English Test Band

4 scores between 400 and 450, indicating an intermediate level of English proficiency; (b) no prior experience with MR devices, ensuring that results would not be influenced by pre-existing familiarity with the technology; and (c) no participation in extracurricular English training programs, ensuring that learning outcomes could be attributed solely to the intervention. Baseline characteristics of the participants were as follows: the age range was 19–21 years, with a mean age of 20.1 and a standard deviation (SD) of 0.8; the gender ratio was approximately 1:1.2, reflecting a relatively balanced distribution; and the mean College English Test Band 4 score was 426.3 with an SD of 31.5. To further verify intergroup homogeneity, a one-way ANOVA was conducted on baseline indicators across the four groups. The results yielded an F value of 0.32 and a p value of 0.86, indicating no statistically significant differences among the groups. This confirmed strong comparability and established a reliable foundation for attributing subsequent learning effects to the experimental intervention.

3.3 Design of the MR instructional intervention

The MR instructional intervention was developed on a standardized technical platform and designed according to the interaction format–scenario–skill adaptability logic, ensuring both reproducibility and specificity. The hardware system was centered on the HoloLens 2 as the primary display device, whose 52° field of view supported an immersive visual experience. Integration with the Leap Motion gesture-tracking module and the Azure speech recognition system enabled precise multimodal data capture and responsive interaction through gestures and speech. The software system, developed using the Unity 2022 engine, incorporated five core functional modules: 1) multimodal perception, 2) virtual character collaboration, 3) spatial scenario environment construction, 4) peer collaboration, and 5) personalized adaptation. The system supported real-time multi-user sharing of virtual spaces and synchronized interaction data, offering stable technical infrastructure for both group-based collaborative tasks and individualized interventions. Figure 1 presents the architecture of the MR-based English education interaction system employed in the study.

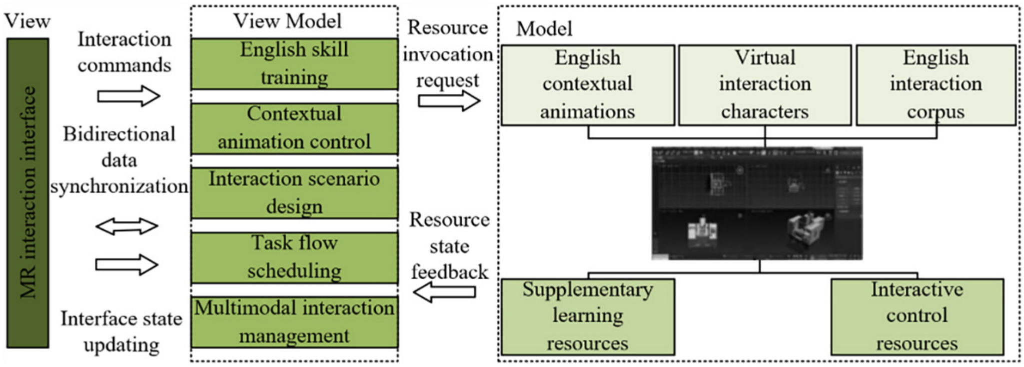


Fig. 1. Architecture of the MR–based English education interaction system

The core adaptability logic of the instructional intervention was operationalized through the precise alignment of interactional formats with skill-specific scenarios. The resulting design is presented in Table 1. All tasks were constructed to align closely with instructional objectives, with contextual environments and multimodal interaction employed to strengthen targeted language skill training. Across the three experimental groups, the instructional content was held constant in terms of difficulty level, thematic coverage, and total training duration. Only the combinations of interactional formats differed. This ensured a fair basis for subsequent comparisons of the relative effectiveness of the MR interactional formats.

Table 1. Adaptability scheme and task examples for MR instructional interaction

Instructional Scenario	Core Interactional Format	Example Task	Instructional Objective
Speaking training	Virtual character collaboration + multimodal perception	A virtual character assumes the role of a human resources officer in an international company, and learners complete an English job interview. Gestures are used to point to résumé modules to present personal experience, spoken responses are provided to interview questions, and the virtual character delivers real-time pronunciation correction and prompts related to eye contact and other professional conventions.	To enhance oral fluency and pronunciation accuracy
Listening training	Spatial scenario immersion + multimodal perception	A virtual airport scenario is presented. Learners move physically through the space while listening to spatialized announcements from different boarding gates, use gesture-based input to record flight information, and complete a situated task involving the identification of the correct gate.	To strengthen listening comprehension in authentic contextual environments
Writing training	Personalized adaptation + virtual-character feedback	A business email is composed. The system recommends templates based on the learner's prior grammatical patterns, and a virtual secretary annotates inappropriate salutations and other issues while offering revision suggestions, enabling real-time refinement.	To improve writing logic and grammatical accuracy
Group classroom task	Peer collaboration + spatial scenario immersion	A virtual business negotiation scenario is enacted. Groups of three learners act as suppliers engaging in negotiation with a virtual buyer. Gesture-based manipulation of price charts is used to support arguments, accompanied by real-time verbal discussion of negotiation strategies.	To improve collaborative communication and language output

3.4 Measurement instruments and data collection

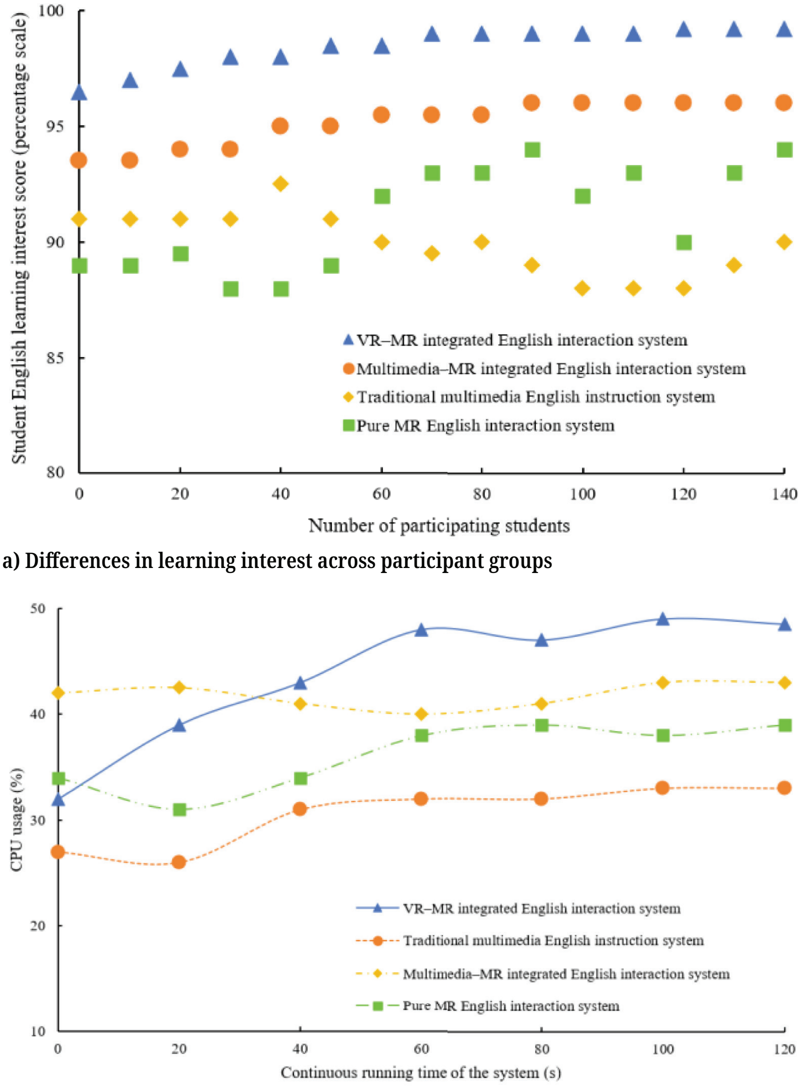
All quantitative instruments employed in this study underwent rigorous reliability and validity testing to ensure the accuracy and consistency of measurement. The language proficiency assessment was adapted from authentic items of the International English Language Testing System and covered five skills—listening, speaking, reading, writing, and translation. Listening, speaking, reading, and writing each accounted for 30 points, and translation accounted for 10 points, yielding a total score of 100 points. Results from the pilot test indicated a Cronbach's α of 0.89. Content validity was evaluated independently by three linguistics professors, and the interrater consistency coefficient of 0.85 indicated strong content validity. The immersion scale was adapted from the game immersion questionnaire and contained 12 items rated on a 7-point Likert scale. Confirmatory factor analysis demonstrated satisfactory model fit, with a χ^2/df ratio of 2.31, root mean square error of approximation of 0.07, and comparative fit index of 0.94. The scale's Cronbach's α coefficient was 0.87. The interaction behavior coding scheme included three core dimensions: interactional format, frequency of meaning negotiation, and length of language output. Meaning negotiation was operationalized as the frequency with which learners produced negotiation moves such as semantic inquiries. Coding was performed independently by two trained researchers, and the inter-coder reliability coefficient of 0.82 met the criteria for acceptable coding reliability. Cognitive load was measured using the Paas scale, consisting of 9 items rated on a 7-point Likert scale. The Cronbach's α coefficient was 0.83, indicating satisfactory reliability.

Data collection was systematically conducted at four key stages to ensure continuity and completeness. During the first week, baseline measurements of language proficiency and learning motivation were administered, followed by two instructional hours of MR device training to ensure all participants had acquired basic interaction skills. Weeks 2 through 16 constituted the intervention phase; at the end of each instructional week, all participants completed the interaction experience scale and cognitive load scale, and the experimental groups additionally submitted reflective learning journals documenting interaction experiences and learning insights. A mid-test was administered in Week 8 to capture intermediate learning outcomes. During Week 16, the post-test of language proficiency was conducted alongside the administration of the technology acceptance questionnaire. Subsequently, ten participants from each group were randomly selected for semi-structured interviews lasting 30 minutes, with the aim of eliciting in-depth subjective perceptions of MR-based interactional instruction. All quantitative data were entered into a database in real time, and qualitative data were transcribed and coded after collection to ensure standardized data management.

3.5 Data analysis methods

A data analytic strategy aligned with the mixed-methods design was employed, with quantitative and qualitative analyses used to corroborate one another to ensure both objectivity and explanatory depth. Quantitative analyses were conducted primarily with SPSS 26.0, and analytical techniques were selected according to the nature of each research question. Repeated-measures ANOVA was used to examine the main effects and the interaction between time and group across the pre-test, mid-test, and post-test language proficiency assessments, thereby determining the persistence of intervention effects and differences among groups. Effect sizes for the ANOVA were reported using partial eta squared (η^2). Between-group effect sizes were reported using Cohen's d, with thresholds defined as $d < 0.5$ for small effects, $0.5 \leq d < 0.8$ for medium effects, and $d \geq 0.8$ for large effects. The PROCESS macro was used to test the mediating role of interaction experience, enabling verification of the chain mechanism linking "MR interactional format → interaction experience → learning outcomes." Hierarchical regression analysis was used to examine the moderating effects of initial language proficiency and cognitive load. The control variable, independent variable, and interaction term were entered stepwise to determine the magnitude and direction of the moderation effects. Statistical significance levels were reported for all quantitative analyses to provide an accurate representation of both practical effect magnitude and statistical reliability. Qualitative data analysis was conducted using NVivo 12 through a thematic coding procedure following a three-step "coding–refinement–verification" process. Open coding was first applied to interview transcripts and reflective journals to identify initial conceptual units. Axial coding was then conducted to organize these units into core thematic categories. Selective coding was subsequently used to construct a thematically saturated framework. The resulting coding result was triangulated with quantitative data to provide deep contextualized explanations for statistical results and to enhance the credibility of the conclusions.

4 RESULTS



a) Differences in learning interest across participant groups

b) Changes in CPU usage

Fig. 2. Experimental results on learning experience and system performance in MR–based English interaction systems

To examine the learning experience gains and the technological performance adaptability of different interaction systems in English language education, the experiment simultaneously monitored students' learning interest and the run-time CPU utilization across four system types, as shown in Figure 2. The interest scores associated with the virtual reality (VR)–MR integrated system remained consistently above 95%. Scores for the pure MR system increased progressively with the growth in participant numbers, eventually stabilizing above 90%. In contrast, the multimedia–MR integrated system and the traditional multimedia system demonstrated markedly lower and more volatile interest scores. These findings indicate that the incorporation of MR technology can produce strong and stable enhancements in students' interest in English learning, with the multi-technology integrated MR system yielding the most pronounced effects. In terms of system performance, the CPU usage of the pure MR system consistently remained within a low and stable range (30%–40%). The VR–MR integrated system exhibited an initial rise in CPU usage, followed by stabilization within a controlled range of 45%–50%. Both MR-based systems demonstrated performance sufficient to support sustained interactional instruction. In contrast, the remaining systems showed either excessively low CPU usage or substantial fluctuation. Taken together, the two types of data indicate that MR technology not only substantially enhances learners' interest and engagement in English education but also demonstrates system performance levels sufficient to support large-scale deployment in instructional settings, providing dual empirical evidence—both experiential and technical—for the feasibility of leveraging MR to enrich interactive English language learning.

To examine the overall impact and temporal stability of MR interaction on English learning outcomes, learning motivation, and perceived immersion, repeated-measures ANOVA and between-group t-tests were conducted. As shown in Table 2, the interaction effect between time and group on overall language proficiency was significant, indicating that the effectiveness of MR interaction increased progressively over time. Post-test scores in the experimental groups improved by 60.1% relative to the pre-test, and a significant difference from the control group had already emerged at the mid-test stage, demonstrating the stability of the intervention effect. In contrast, the control group exhibited only a 34.6% improvement in the post-test, with no notable progress at mid-test. Regarding psychological experience, both learning motivation and immersion were significantly higher in the experimental groups than in the control group at post-test, with immersion yielding an effect size of 3.86, indicating an exceptionally strong effect. Overall, the findings show that MR interaction not only substantially enhances English learning outcomes but also simultaneously improves learning motivation and immersion, with effects that remain robust over time. These results offer core empirical support for the integration of MR technology into English language education.

Table 2. Overall impact of MR interaction on learning outcomes and psychological experience

Measure	Group	Pre-Test (M ± SD)	Mid-Test (M ± SD)	Post-Test (M ± SD)	Time × Group Interaction (F/η^2)	Post-Test Group Difference (t/d)
Overall language proficiency	Control	48.6 ± 5.9	55.2 ± 6.4	65.4 ± 7.2	28.63/0.45 ($p < 0.001$)	8.92/1.89 ($p < 0.001$)
	Experimental	49.1 ± 6.1	62.8 ± 6.7	78.6 ± 6.3		
Learning motivation (7-point scale)	Control	3.12 ± 0.60	3.10 ± 0.58	3.15 ± 0.62	–	8.72/1.92 ($p < 0.001$)
	Experimental	3.08 ± 0.59	3.85 ± 0.55	4.23 ± 0.51	–	
Immersion (7-point scale)	Control	3.18 ± 0.75	3.19 ± 0.76	3.21 ± 0.78	–	17.35/3.86 ($p < 0.001$)
	Experimental	3.20 ± 0.73	4.92 ± 0.68	5.87 ± 0.64	–	

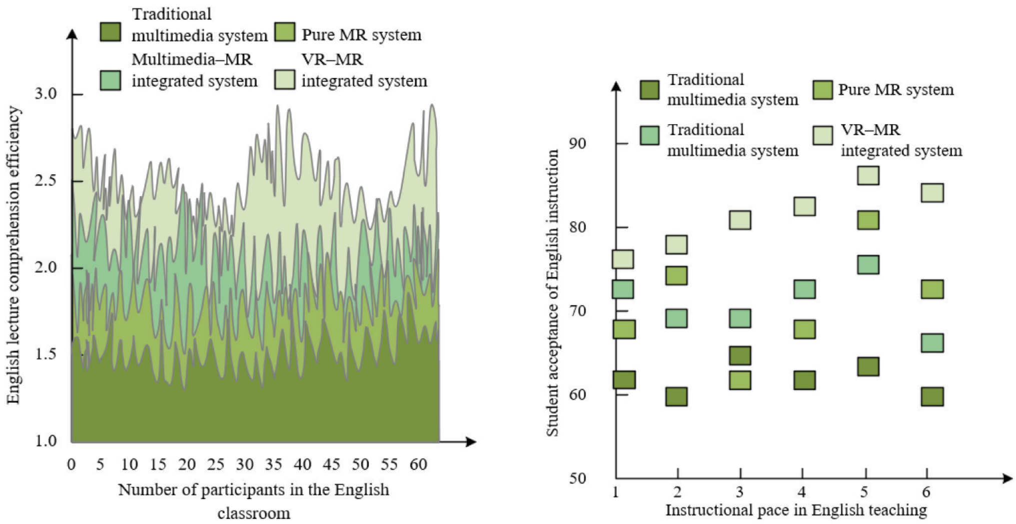
To examine the adaptability patterns between MR interactional formats and specific English language skills, effect-size analyses and significance testing were conducted across groups. As shown in Table 3, all experimental groups demonstrated substantially greater improvement across skills when compared with the control group, and the results exhibited clear adaptability characteristics.

For speaking skills, experimental group 1 produced the most pronounced effect, with an improvement of 41.2% and an effect size approaching 2.0, indicating an exceptionally strong impact. Qualitative evidence suggested that the low-anxiety interaction environment facilitated by virtual characters constituted a key contributing factor. For listening skills, experimental group 2 generated the greatest enhancement, reflected by a 38.5% improvement and an effect size of 1.85. These findings corroborate the strengthening effect of spatialized contextual environments on authentic listening input. Regarding writing and translation, experimental group 3 exhibited the most substantial gains. Writing improvement reached 39.1%, with an effect size of 2.03, indicating that personalized feedback offered precise optimization of written expression. Although overall gains in reading were comparatively moderate, experimental group 3 still produced the highest improvement (30.4%). Collectively, these results demonstrate that distinct MR interactional formats exhibit clear adaptation relationships with different English skills, providing direct empirical support for the design of skill-targeted MR instructional schemes.

Table 3. Comparative improvements in English skills across MR interaction formats

Skill Dimension	Control Group	Experimental Group 1 (Virtual Character + Multimodality)	Experimental Group 2 (Spatial Scenario + Peer Collaboration)	Experimental Group 3 (Personalized Adaptation + Multimodality)
Speaking	Improved by 12.3%	Improved by 41.2%	Improved by 35.7%	Improved by 38.9%
	d = 0.42 (small)	d = 1.98 (large, p < 0.001)	d = 1.65 (large, p < 0.001)	d = 1.82 (large, p < 0.001)
Listening	Improved by 14.5%	Improved by 36.8%	Improved by 38.5%	Improved by 37.2%
	d = 0.48 (small)	d = 1.72 (large, p < 0.001)	d = 1.85 (large, p < 0.001)	d = 1.76 (large, p < 0.001)
Writing	Improved by 16.7%	Improved by 32.4%	Improved by 28.6%	Improved by 39.1%
	d = 0.53 (small)	d = 1.43 (medium, p < 0.01)	d = 1.28 (medium, p < 0.01)	d = 2.03 (large, p < 0.001)
Reading	Improved by 18.2%	Improved by 25.3%	Improved by 27.8%	Improved by 30.4%
	d = 0.59 (small)	d = 0.92 (medium, p < 0.05)	d = 1.05 (medium, p < 0.05)	d = 1.18 (medium, p < 0.01)
Translation	Improved by 15.6%	Improved by 33.7%	Improved by 34.2%	Improved by 35.8%
	d = 0.51 (small)	d = 1.49 (medium, p < 0.01)	d = 1.52 (medium, p < 0.01)	d = 1.63 (large, p < 0.001)

To clarify the differences in effectiveness among various interactive systems in accommodating different class sizes and pacing in English instruction, this experiment examined the lecture comprehension efficiency and student acceptance of four types of systems, as shown in Figure 3. For both the pure MR system and the VR–MR integrated system, listening efficiency consistently remained above 1.5, even as the number of classroom participants increased, with minimal fluctuations. In contrast, the traditional multimedia system exhibited lower overall efficiency, accompanied by more pronounced variation as participant numbers increased. These patterns indicate that the incorporation of MR technology enhances the stability of listening efficiency across diverse classroom scales. In terms of student acceptance, the VR–MR integrated system and the multimedia–MR integrated system both maintained acceptance levels above 70 throughout the instructional pacing continuum. The pure MR system sustained acceptance scores above 60, with no downward trend as instructional speed increased. Conversely, acceptance levels for the traditional multimedia system remained low and declined further as the instructional pace accelerated. These findings demonstrate that MR-based instructional systems exhibit superior adaptability to variations in instructional pacing and more effectively sustain learner acceptance. Taken together, the results indicate that MR technology significantly strengthens both listening efficiency and student acceptance across a range of instructional conditions. The adaptability of MR-based systems to diverse instructional scenarios provides empirical support for the flexible implementation of MR in interactive English education.



a) Changes in lecture comprehension efficiency with increasing class size

b) Changes in student acceptance with increasing instructional pace

Fig. 3. Experimental results on lecture comprehension efficiency and student acceptance across different interaction systems

To elucidate the mechanism through which MR interaction influences English learning outcomes, mediation analysis was conducted using the PROCESS macro. As shown in Table 4, the total effect of MR interaction on learning outcomes was significant ($\beta = 0.62$). When the mediators—immersion, depth of meaning, negotiation, and engagement—were included in the model, the direct effect of MR interaction declined to 0.13 and became nonsignificant. The total indirect effect was significant ($\beta = 0.49$) and accounted for 79.0% of the total effect, indicating a full mediation pattern. Decomposition of the mediation pathways revealed that immersion served as the strongest mediator ($\beta = 0.28$), accounting for 57.1% of the total indirect effect. The depth of meaning negotiation constituted the second strongest mediator, with a coefficient of 0.21 and a contribution of 42.9%. Engagement exhibited a comparatively weaker mediating effect ($\beta = 0.08$), contributing 16.3% of the total indirect effect. These findings indicate that MR interaction does not exert its influence on learning outcomes directly; rather, its effects are transmitted through enhancements in immersion and the deepening of meaning negotiation, with immersion representing the primary mediating mechanism and meaning negotiation providing essential supplementary support. This mechanism offers a clear instructional implication: MR-based pedagogical design should prioritize immersive environment construction and the design of meaning-negotiation scenarios.

Table 4. Mediation analysis of interaction experience between MR interaction and learning outcomes

Effect Type	Regression Coefficient (β)	Standard Error (SE)	t-Value	p-Value	Proportion Mediated
Total effect of MR interaction on learning outcomes	0.62	0.07	8.86	<0.001	–
Direct effect of MR interaction on learning outcomes	0.13	0.08	1.63	0.12	–
Mediation effect of immersion	0.28	0.05	5.60	<0.001	57.1%
Mediation effect of depth of meaning negotiation	0.21	0.06	3.50	<0.001	42.9%
Mediation effect of engagement	0.08	0.04	2.00	0.04	16.3%
Total mediation effect	0.49	0.06	8.17	<0.001	79.0%

Note: The proportion mediated was calculated as (coefficient of each mediation path/coefficient of total mediation effect) × 100%. Mediation effects were examined using PROCESS Macro Model 4.

To identify the applicability boundaries and optimal conditions for the effectiveness of MR interaction, hierarchical regression analysis was conducted to examine the moderating roles of language proficiency and cognitive load. As shown in Table 5, language proficiency demonstrated a significant moderating effect, with the interaction term yielding a coefficient of 0.23 ($p < 0.01$). Learners with lower or intermediate proficiency levels exhibited a 38.6% improvement, accompanied by an effect size of 2.15—substantially exceeding the 27.3% improvement and effect size of 1.28 observed among higher-proficiency learners. These results indicate that MR interaction contributes meaningfully to reducing disparities in language proficiency. Cognitive load also exhibited a significant moderating effect, reflected by a negative interaction coefficient ($\beta = -0.19$, $p < 0.01$). The low–cognitive-load group demonstrated a 39.2% improvement and an effect size of 2.21, in contrast to the 28.5% improvement and effect size of 1.32 observed in the high–cognitive-load group. This pattern confirms that overly complex interaction design attenuates the effectiveness of MR-based instruction. Taken together, the moderating effects of language proficiency and cognitive load illustrate that MR interaction outcomes are jointly shaped by learners' foundational abilities and the complexity of interactional design. The findings further indicate that learners with low or intermediate proficiency constitute the primary beneficiaries of MR-based instruction and that interactional design should adhere to a "low-load, high-adaptability" principle, providing essential guidance for the precise implementation of MR in English education.

Table 5. Moderating effects of language proficiency and cognitive load on the impact of MR interaction

Moderating Variable	Moderation Level	Improvement Percentage	Post-Test Score (M ± SD)	Effect Size (d)	Interaction Term Coefficient (β/p)
Language proficiency	High level	27.3%	82.5 ± 5.8	1.28	0.23/ < 0.01
	Medium-low level	38.6%	76.3 ± 6.5	2.15	
Cognitive load	High load	28.5%	74.2 ± 6.7	1.32	–0.19/ < 0.01
	Low load	39.2%	79.1 ± 6.1	2.21	
Control group (baseline)	–	15.2%	65.4 ± 7.2	0.53	–

Note: Improvement percentage = (post- – pre-test score)/pre-test score × 100%. Interaction term coefficients were obtained through hierarchical regression analysis.

5 CONCLUSION

This study was conducted with the aim of addressing three structural limitations of traditional English education: the singularity of interactional forms, the disconnection between instructional scenarios and authentic language use, and the inefficiency of skill transfer. Focusing on the scenario adaptability and skill specificity of MR interactional formats and drawing on SLA theory and multimodal interaction theory, a 16-week randomized controlled and mixed-methods intervention was implemented with 90 university students majoring in subjects other than English. One traditional multimedia control group and three experimental groups—each receiving a distinct MR interactional format (virtual character collaboration, spatial scenario immersion, or personalized adaptation)—were evaluated using language proficiency assessments, interaction behavior coding, psychological scales, and semi-structured interviews. The principal findings revealed that the experimental groups using MR achieved significantly greater improvements in overall language proficiency than the control group, with effects that strengthened over time. Distinct adaptability patterns emerged across language skills: virtual character collaboration produced the strongest gains in speaking; spatial scenario immersion yielded the greatest enhancements in listening; and personalized adaptation generated the most substantial improvements in writing. Immersion and the depth of meaning negotiation were shown to function as full mediators between MR interaction and learning outcomes. Learners with lower or intermediate initial proficiency benefited more markedly from MR-based instruction, and low-cognitive-load interaction design amplified the effectiveness of the intervention. The core contribution of this study lies in the development of a three-dimensional interaction format–scenario–skill adaptability model, which provides a mechanism-based explanation for the application of multimodal interaction theory within the domain of SLA. From a practical perspective, the study establishes a "precision adaptation" design paradigm for technology-enhanced language education, offering actionable principles for the development and classroom deployment of MR instructional systems and reducing the risk of unfocused or inefficient technological integration.

6 ACKNOWLEDGEMENTS

This paper was funded by the Science Research Project of the Hebei Education Department, "A Study on the Interdisciplinary Cooperative Teaching Model of College English Based on ESP Transformation" (Grant No.: SQ2024122).

7 REFERENCES

[1] J. Higgins, "Virtual bridges to global competence: Cultivating undergraduates' English communication and intercultural skills through online exchange," *Journal of International Students*, vol. 14, no. 4, pp. 862–882, 2024. https://doi.org/10.32674/jis.v14i4.6545

[2] M. K. Ibrahim and S. Glushkov, "Boosting adolescent English proficiency: Psycholinguistic benefits of mobile platforms," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 19, pp. 83–105, 2024. https://doi.org/10.3991/ijim.v18i19.48393

[3] F. Chen, "Study on the multivariant interactive teaching modes of college English under the information technology environment," *Informatica*, vol. 43, no. 3, pp. 343–347, 2019. https://doi.org/10.31449/inf.v43i3.2807

[4] L. Holubnycha *et al.*, "AI-powered teaching: Literature review of ChatGPT's impact on university educators," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 15, pp. 41–65, 2025. https://doi.org/10.3991/ijim.v19i15.54275

[5] D. Tatzl, "Case meetings for teaching English for specific academic purposes in a tertiary aeronautical engineering programme," *Innovation in Language Learning and Teaching*, vol. 9, no. 3, pp. 191–217, 2015. https://doi.org/10.1080/17501229.2014.904871

[6] S. Turkan, "In-service teachers' reasoning about scenarios of teaching mathematics to English language learners," *The Mathematics Enthusiast*, vol. 13, no. 1, pp. 130–148, 2016. https://doi.org/10.54870/1551-3440.1369

[7] A. S. Elsaid Mohammed and H. S. M. Nur, "Needs analysis in English for academic purposes: The case of teaching assistants at the University of Khartoum," *How*, vol. 25, no. 2, pp. 49–68, 2018. https://doi.org/10.19183/how.25.2.409

[8] M. J. Mante-Estacio, P. Nino Valdez, and D. Pulido, "Effective teaching of the macroskills: Reflections from Filipino teachers of English," *Reflective Practice*, vol. 19, no. 6, pp. 844–854, 2018. https://doi.org/10.1080/14623943.2018.1539662

[9] M. Jamil, H. Hadiyanto, and R. Sanjaya, "Evaluation of mobile mixed reality simulator (M2RSi) using hybrid technology acceptance model (HTAM) in disaster mitigation simulation," *International Journal of Safety and Security Engineering*, vol. 14, no. 2, pp. 563–572, 2024. https://doi.org/10.18280/ijsse.140223

[10] X. Chen and W. Zhang, "Research on design and evaluation system of traditional Chinese medicine teaching system based on MR mixed-reality model," *Journal of Computational Methods in Sciences and Engineering*, 2025. https://doi.org/10.1177/14727978251385186

[11] J. Fernando and T. Oktavia, "Evaluation of the utilization of augmented reality technology to increase interaction satisfaction in remote assistance," *Journal Européen des Systèmes Automatisés*, vol. 58, no. 4, pp. 755–761, 2025. https://doi.org/10.18280/jesa.580409

[12] S. Liu, P. Toreini, and A. Maedche, "Mixed reality learning systems with head-mounted displays in higher education: A systematic review," *Technology, Knowledge and Learning*, pp. 1–35, 2025. https://doi.org/10.1007/s10758-025-09912-z

[13] K. Mäenpää, H. Järvenoja, J. Peltonen, and K. Pyhältö, "Nursing students' motivation regulation strategies in blended learning: A qualitative study," *Nursing & Health Sciences*, vol. 22, no. 3, pp. 602–611, 2020. https://doi.org/10.1111/nhs.12702

[14] W. Fang, T. Zhang, L. Chen, Y. Cao, H. Hu, and J. Bi, "HiMAR: A hierarchical context-aware mobile AR assembly for large-scale cable products," *International Journal of Production Research*, pp. 1–22, 2025. https://doi.org/10.1080/00207543.2025.2568740

[15] W. Fang, T. Zhang, Z. Wang, and J. Ding, "A multi-modal context-aware sequence stage validation for human-centric AR assembly," *Computers & Industrial Engineering*, vol. 194, p. 110355, 2024. https://doi.org/10.1016/j.cie.2024.110355

[16] M. Araki, A. Kouzawa, and K. Tachibana, "Proposal of a multimodal interaction description language for various interactive agents," *IEICE Transactions on Information and Systems*, vol. 88, no. 11, pp. 2469–2476, 2005. https://doi.org/10.1093/ietisy/e88-d.11.2469

8 AUTHORS

Jingjie Zhang graduated from Northeast Forestry University in 2015 and is currently affiliated with the Department of Basic Courses, Hebei University of Water Resources and Electric Engineering, Cangzhou 061000, China. Research interests include college English teaching practice and English language teaching methodologies (E-mail: czzhangjingjie@163.com).

Youzi Liang graduated from Nankai University in 2017 and is currently affiliated with the College of Business Administration, Hebei Vocational University of Industry and Technology, Shijiazhuang 050091, China. Research interests include English language teaching, international exchange and education, and home economics education (E-mail: liangyouzi819@163.com).