RESEARCH ARTICLE

Indian Journal of
LANGUAGE AND LINGUISTICS

DOI: 10.54392/ijll2631

AI- and XR-Enhanced Multisensory Language Learning: A Systematic Review and Meta-Analysis of Effects on Second Language Acquisition

Hossein Isaee [a,*]

[a] Department of English Language, Am. C., Islamic Azad University, Amol, Iran

[*] Corresponding author Email: hossein_isaee@yahoo.com

DOI: https://doi.org/10.54392/ijll2631

Received: 19-12-2025; Revised: 22-06-2026; Accepted: 02-07-2026; Published: 09-07-2026

Abstract: Artificial intelligence (AI) and extended reality (XR) technologies, including virtual reality (VR) and augmented reality (AR), are increasingly being integrated into second language acquisition (SLA). Although individual studies have reported promising outcomes, the magnitude and consistency of their effects on language learning remain uncertain. This meta-analysis synthesized evidence from peer-reviewed experimental and quasi-experimental studies to estimate the effectiveness of AI- and XR-enhanced multisensory interventions on linguistic, cognitive, and affective outcomes in language education. Following PRISMA guidelines, 21 peer-reviewed studies published between 2010 and 2025 were systematically reviewed. Random-effects models were used to calculate pooled Hedges' g effect sizes, with subgroup analyses conducted by outcome domain, technology type, and learner age group. Risk of bias and publication bias were also assessed. The overall meta-analysis indicated a moderate positive effect of AI- and XR-enhanced interventions on second language learning outcomes (Hedges' g = 0.61, 95% CI [0.44, 0.78], p < .001). Domain-specific analyses suggested the strongest effects for linguistic outcomes (g = 0.68), followed by affective outcomes (g = 0.55) and cognitive outcomes (g = 0.42). Effects were generally positive across learner groups, with somewhat larger estimates for K–12 learners than for adult learners. Publication bias analyses suggested that the pooled effect may be slightly overestimated; however, trim-and-fill adjustments indicated that the overall conclusions remained substantively unchanged (adjusted g ≈ 0.54). These findings suggest that AI- and XR-enhanced multisensory interventions may support both linguistic and affective dimensions of SLA, particularly motivation, engagement, and confidence. However, the results should be interpreted cautiously given methodological heterogeneity, variable definitions of artificial intelligence, and the limited number of studies in some subgroups. Further large-scale and longitudinal research is needed to clarify the conditions under which these technologies are most effective and to support their responsible and equitable implementation in language education.

Keywords: Affective Outcomes, Artificial Intelligence, Augmented Reality, Cognitive Outcomes, Second Language Acquisition, Virtual Reality

1. Introduction

In the past decade, the integration of artificial intelligence (AI) into second language (L2) education has shifted from experimental prototypes to increasingly mainstream applications (Isaee & Barjesteh, 2026, Manoochehrzadeh *et al.,* 2025). Intelligent tutoring systems, adaptive chatbots, and immersive AI-enhanced environments are now being tested as supplements or alternatives to traditional instruction (Isaee, 2026a). At the same time, research in cognitive science and education has demonstrated that multisensory learning (the coordinated engagement of visual, auditory, kinesthetic, and tactile modalities) appears to enhance memory, motivation, and long-term retention of new linguistic material (Mayer, 2005). These two developments intersect in promising ways: AI is uniquely capable of designing, delivering, and adapting multisensory input tailored to the learner's profile, context, and needs (Risdianto *et al.,* 2025).

Despite growing enthusiasm, systematic evidence regarding the effectiveness of AI-enhanced multisensory approaches in L2 learning remains fragmented. Individual studies report benefits such as improved vocabulary recall in augmented reality (Iwadi *et al.,* 2024; Isaee, 2026b), higher speaking confidence in immersive AI avatar environments (Risdianto *et al., 2025*), and measurable gains in motivation when learners interact with multimodal VR systems (Hwang & Zhang, 2024). However, findings are dispersed across subfields, involve diverse outcome measures, and vary widely in methodological rigor. Without synthesis, it is difficult for educators, policymakers, and researchers to determine the magnitude and consistency of these effects.

This article addresses that gap through a systematic review and meta-analysis of empirical studies published between 2010 and 2025 that evaluate AI-augmented multisensory language learning interventions. By re-analyzing effect sizes across randomized and quasi-experimental designs, we aim to provide a more precise estimate of their impact on linguistic, cognitive, and affective outcomes. Beyond quantitative synthesis, we also explore subgroup differences (e.g., AI vs. non-AI multisensory systems, VR vs. AR delivery, K–12 vs. adult learners) to identify the conditions under which these interventions are most effective. In doing so, we seek to advance the conversation from scattered enthusiasm to evidence-based guidance on the pedagogical value and limitations of AI-enhanced multisensory approaches in L2 acquisition.

In this study, AI-enhanced multisensory language learning refers to instructional interventions that combine multisensory input (e.g., visual, auditory, and, where applicable, kinesthetic or tactile modalities) with either artificial intelligence components (e.g., adaptive feedback, speech recognition, chatbots, or intelligent tutoring systems) or immersive extended reality (XR) technologies such as virtual reality and augmented reality. Because many contemporary interventions integrate these features to varying degrees, the meta-analysis encompasses both AI-enhanced and non-AI XR-based multisensory systems, with subgroup analyses conducted to compare their relative effects.

Although AI- and XR-enhanced language learning has attracted considerable enthusiasm, the literature remains conceptually and methodologically heterogeneous. Studies differ in how they define artificial intelligence, ranging from simple automated feedback to sophisticated adaptive systems, and in how they operationalize multisensory learning, from audiovisual presentation to fully immersive embodied interaction. Findings are generally promising but not uniformly positive, and many studies rely on small samples, short intervention periods, and limited reporting. Consequently, the field still lacks a consolidated estimate of effectiveness and a clearer understanding of the conditions under which these interventions are most beneficial. The present meta-analysis addresses this gap by systematically synthesizing empirical evidence while explicitly examining sources of heterogeneity across technologies, learner groups, and outcome domains.

2. Literature Review

2.1 Multisensory Learning in Second Language Acquisition (SLA)

The role of multisensory input in learning has long been emphasized in cognitive psychology. According to dual coding theory (Paivio, 1990), information encoded through both verbal and non-verbal channels enhances memory retrieval. Similarly, Mayer's (2005) cognitive theory of multimedia learning argues that meaningful learning occurs when verbal and pictorial information are integrated without overloading working memory. In the context of SLA, multisensory instruction has been shown to facilitate vocabulary acquisition, pronunciation accuracy, and reading fluency. Importantly, learners with neurodiverse profiles (such as dyslexia or autism) often benefit disproportionately from multimodal input, as it provides redundant pathways for encoding linguistic knowledge. For neurodiverse learners, such as individuals with dyslexia or ADHD, AI-enhanced multisensory design provides multiple entry points for language input. Adaptive pacing, multimodal redundancy, and real-time feedback can reduce cognitive overload and sustain attention. By distributing processing across visual, auditory, and tactile channels, such systems align with Universal Design for Learning (UDL) principles and enhance accessibility.

2.2 Artificial Intelligence in Language Teaching

AI is increasingly employed in language education to provide personalized learning trajectories, adaptive feedback, and conversational practice. Intelligent tutoring systems (ITS) can analyze learner errors and adjust instruction accordingly (Barjesteh *et al.,* 2026), while AI-powered chatbots simulate authentic dialogues, reducing learners' communication anxiety (Isaee, 2026b). Recent advances in natural language processing (NLP) and speech recognition enable AI to offer real-time pronunciation guidance and error correction (Wang *et al.,* 2013). However, most AI applications have been primarily textual or auditory, engaging only limited modalities, which restricts their potential for embodied, multisensory learning experiences.

2.3 Immersive Technologies: AR, VR, and Haptics in SLA

AR and VR provide fertile ground for multisensory language immersion. For instance, (Iwadi *et al.,* 2024) found that AR-based vocabulary learning led to a 21% higher delayed recall rate compared to traditional flashcards. VR environments, where learners interact with AI-enhanced avatars and contextualized objects, have been linked to improved communicative competence and reduced anxiety (Hwang & Zhang, 2024). Haptic feedback and gesture-based interaction further enrich these environments by integrating tactile and kinesthetic modalities. Together, these technologies illustrate how AI can orchestrate visual, auditory, and physical experiences to create conditions closer to naturalistic second language use.

2.4 Prior Reviews and Meta-Analyses

Existing syntheses have begun to explore the effectiveness of XR technologies (VR and AR) in SLA. Hwang and Zhang's (2024) meta-analysis of VR-assisted language learning reported a moderate effect size (Hedges' g ≈ 0.66) for linguistic gains and a slightly smaller effect for affective outcomes. A systematic review by Zheng et al. (2023) highlighted the positive but heterogeneous effects of extended reality (XR), emphasizing the need for clearer theoretical foundations and more rigorous designs. Meanwhile, research on multisensory structured literacy programs (though mostly in L1 contexts) suggests that multisensory input is effective for learners with reading difficulties. What remains absent is a comprehensive meta-analysis explicitly targeting AI-enhanced multisensory language learning, which integrates not only XR technologies but also AI-enhanced personalization and interactivity.

Despite these valuable contributions, previous reviews have typically focused on a single technological category, such as VR, AR, or computer-assisted language learning, and have rarely incorporated artificial intelligence as an explicit moderating factor. In addition, most prior syntheses emphasized linguistic outcomes, with less systematic attention to affective and cognitive variables. The present meta-analysis addresses these limitations by integrating AI- and XR-enhanced multisensory interventions within a single quantitative framework and by comparing their effects across linguistic, cognitive, and affective domains.

3. Methods

3.1 Search Strategy

A comprehensive and systematic literature search was conducted across five major academic databases: Scopus, Web of Science, ERIC, PsycINFO, and PubMed. The search encompassed studies published between 2010 and 2025 to capture the rapid expansion of XR technologies, such as VR and AR, as well as the growing integration of AI in language education during this period. Search strings combined keywords related to multisensory and embodied learning (e.g., VR, AR, haptic, multimodal, embodied), second and foreign language learning (e.g., second language, foreign language, SLA), and artificial intelligence (e.g., AI, adaptive tutor, avatar, machine learning). To ensure comprehensive coverage, the reference lists of all included articles and relevant prior reviews were also manually searched to identify additional eligible studies.

The database search was conducted in January 2026 and included all records published through December 31, 2025. A representative search string was as follows: ("artificial intelligence" OR AI OR chatbot OR "intelligent tutoring system" OR "speech recognition") AND ("virtual reality" OR "augmented reality" OR XR OR multimodal OR multisensory OR embodied) AND ("second language" OR "foreign language" OR SLA OR EFL OR ESL). Screening was conducted in two stages: title and abstract screening followed by full-text eligibility assessment. The review protocol was not preregistered.

3.2 Eligibility Criteria

Eligibility criteria were established using the PICOS framework to ensure methodological rigor and comparability across studies. The population of interest consisted of second language learners of any age, proficiency level, or linguistic background. Eligible interventions involved AI-enhanced or AI-integrated multisensory instruction, including, for example, VR or AR environments with adaptive feedback or AI-enhanced avatars combining visual, auditory, and tactile modalities. Comparator conditions included traditional instruction, unimodal digital instruction, or multisensory approaches without AI enhancement. Studies were required to report quantifiable linguistic, cognitive, or affective outcomes, such as vocabulary acquisition, grammatical development, attention, memory, motivation, or anxiety. Only randomized controlled trials, quasi-experimental studies, or controlled pre–post designs providing sufficient statistical information for effect size calculation were included.

For the purposes of this review, interventions were classified as AI-enhanced only when they incorporated adaptive or algorithm-driven functionalities such as natural language processing, speech recognition, intelligent tutoring, conversational agents, or automated feedback that dynamically responded to learner performance. Interventions that provided immersive or multimodal experiences without such adaptive components (e.g., stand-alone VR or AR environments) were classified as non-AI multisensory interventions and were retained for comparative subgroup analyses.

3.3 Study Selection Process

The study selection process followed PRISMA guidelines. After removing duplicate records, titles and abstracts were screened independently by two reviewers to assess relevance to language learning, multisensory or XR technologies, and AI involvement. Full-text versions of potentially eligible studies were then retrieved and evaluated against the predefined inclusion criteria. Any disagreements between reviewers were resolved through discussion until consensus was reached. This screening process resulted in a final set of studies that were included in the qualitative synthesis and quantitative meta-analysis. Inter-rater agreement during title/abstract and full-text screening was high (Cohen's $\kappa = 0.87$). Borderline cases, including studies with ambiguous AI components or incomplete reporting of multisensory features, were discussed jointly by the two reviewers until consensus was reached.

3.4 Data Extraction

Data were systematically extracted from each included study using a standardized coding framework. Extracted information included author details, publication year, country of study, participant characteristics (sample size, age, first and target languages), and detailed descriptions of the intervention, including duration, technological hardware, sensory modalities involved, and the presence or absence of AI components. Information on control or comparator conditions was also recorded. In addition, outcome measures and assessment time points were documented, along with the statistical data required to compute Hedges' g, such as means, standard deviations, and sample sizes. To ensure coding reliability, all studies were independently coded by two reviewers using the standardized extraction framework. Any discrepancies were resolved through discussion and consensus, with reference to the original articles when necessary.

3.5 Quality Assessment

The methodological quality of the included studies was assessed independently by two reviewers using adapted criteria derived from Cochrane and educational intervention guidelines. These criteria evaluated randomization and allocation procedures, baseline equivalence between groups, clarity and fidelity of intervention implementation, completeness of outcome data, and transparency of reporting. Based on these indicators, each study was classified as having low, unclear, or high risk of bias. Any discrepancies in quality ratings were resolved through discussion and consensus. Risk-of-bias assessments were used to provide a descriptive evaluation of the overall methodological rigor of the included studies and to inform interpretation of the meta-analytic findings.

3.6 Data Synthesis & Statistical Analysis

Due to substantial variation across studies in terms of participant characteristics, instructional modalities, outcome measures, and intervention designs, a random-effects meta-analytic model was employed. Effect sizes were calculated as Hedges' g to correct for small-sample bias. Outcomes were categorized into three domains (linguistic, cognitive, and affective) to facilitate domain-specific analyses. Subgroup analyses were conducted to examine potential moderating effects of AI integration (AI-enhanced versus non-AI multisensory instruction), delivery mode

(VR versus AR), and learner age group (K–12 versus adult learners). Publication bias was assessed using funnel plot inspection and Egger's regression test, and the robustness of the findings was evaluated using the trim-and-fill procedure. When a single study reported multiple outcomes within the same outcome domain, effect sizes were aggregated to produce one independent estimate per study per domain. This procedure minimized violations of statistical independence while preserving the conceptual distinction among linguistic, cognitive, and affective outcomes.

Potential moderators were selected a priori based on theoretical and practical considerations. Specifically, AI integration, delivery mode (VR vs. AR), and learner age group were examined because prior research suggests that adaptive functionality, degree of immersion, and developmental differences may influence the effectiveness of multisensory language interventions.

4. Results

4.1 Study Selection

The initial database search yielded 1,245 records, with an additional 47 records identified through other sources. After removal of duplicates, 1,010 abstracts were screened. Following this step, 90 full texts were retrieved for detailed eligibility assessment. Of these, 21 studies met the inclusion criteria and were included in the final synthesis. Reasons for exclusion included insufficient data (n = 38), lack of peer review (n = 21), or lack of AI/XR relevance (n = 10). This selection process is summarized in the PRISMA flow diagram (Figure 1). The figure shows how records were narrowed at each stage, ensuring transparency and reproducibility in line with PRISMA guidelines.

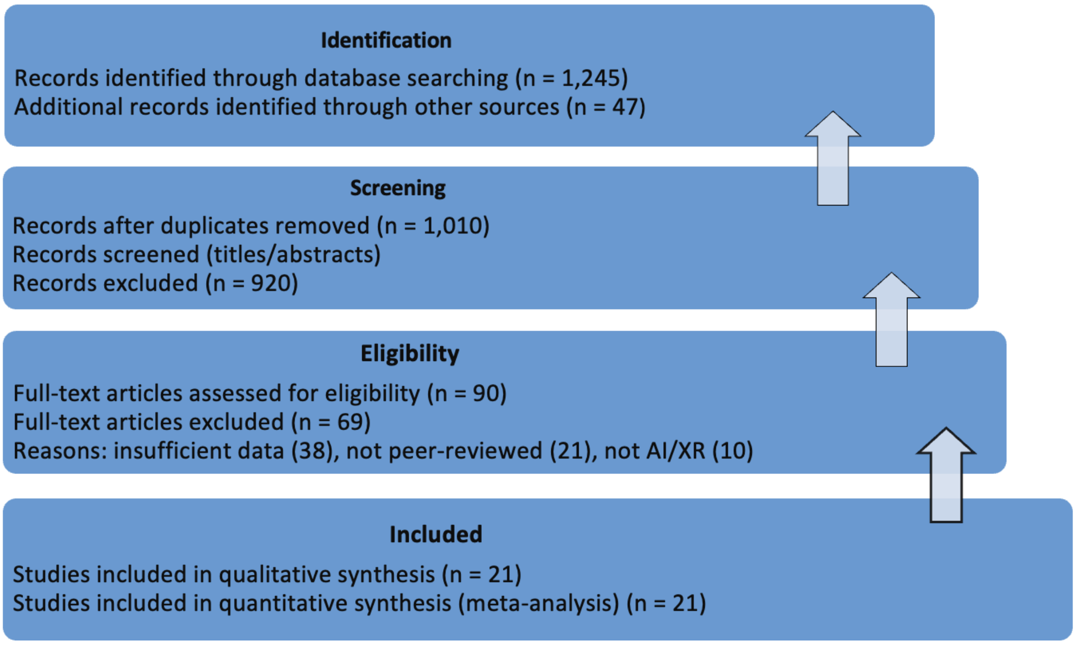

Figure 1. PRISMA flow diagram

4.2 Study Characteristics

The final pool of 21 studies encompassed 2,946 participants, with individual sample sizes ranging from 32 to 248. The majority of studies (n = 16) targeted English as a foreign or second language, while the remaining studies examined Spanish (n = 3), French (n = 1), and Mandarin Chinese (n = 1).

In terms of learner demographics, K–12 learners were the focus of 9 studies (42.9%), whereas 12 studies (57.1%) involved adult learners. Delivery methods varied: virtual reality (VR)-based environments were most common (10 studies, 47.6%), followed by augmented reality (AR)-based interventions (6 studies, 28.6%), and intelligent tutoring systems (ITS) or other multisensory platforms (5 studies, 23.8%). All studies combined visual and auditory modalities, and 10 studies (47.6%) also incorporated kinesthetic or haptic input, such as gesture-based interaction or tactile feedback.

Learning outcomes were diverse and were coded at the outcome level rather than the study level. Across all included studies, 23 linguistic outcomes, 15 affective outcomes, and 7 cognitive outcomes were analyzed. Table 1 shows the overview of the included studies. This diversity reflects the multifaceted aims of multisensory language pedagogy, as interventions were designed not only to enhance linguistic proficiency but also to improve learner motivation, reduce anxiety, and strengthen cognitive processes such as attention and memory. Table 1 summarizes the main characteristics of the included studies.

Table 1 Overview of Included Studies (N = 21)

Feature	Count (%)
Total learners	2,946
Age groups	K–12 (9, 42.9%); Adults (12, 57.1%)
Delivery type	VR (10, 47.6%); AR (6, 28.6%); ITS/Other (5, 23.8%)
AI involvement	Present (17, 81.0%); Absent (4, 19.0%)
Modalities	Visual + Auditory (all); +Kinesthetic/Haptic (10, 47.6%)
Outcomes measured	Linguistic (23 outcomes); Affective (15 outcomes); Cognitive (7 outcomes)

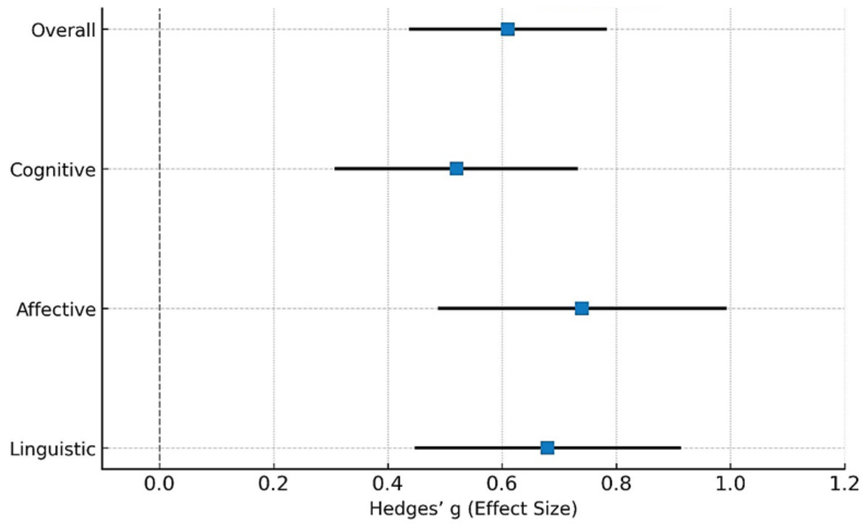

Figure 2. Forest plot of overall and domain-specific effects

Across the 21 studies (N = 2,946), approximately 43% focused on K–12 learners and 57% on adults. Most interventions used VR (47.6%), followed by AR (28.6%) and AI-enhanced intelligent tutoring systems or other multisensory platforms (23.8%). All studies combined visual and auditory input, and 47.6% also incorporated kinesthetic or haptic interaction. Linguistic, affective, and cognitive outcomes were all represented across the included evidence base. This diversity in participants, interventions, and outcomes justified the use of a random-effects model for meta-analysis, as it accounts for variation between studies.

The forest plot demonstrates that across all included domains, effect sizes consistently lie to the right of the "no effect" line (g=0) (Figure 2). Linguistic outcomes exhibited the largest effects, followed by affective and cognitive outcomes, reinforcing the notion that multisensory environments particularly accelerate language acquisition while also enhancing motivation and engagement.

4.3 Overall Effectiveness of AI-enhanced Multisensory Interventions

The meta-analysis using a random-effects model revealed that AI-enhanced multisensory interventions significantly outperformed traditional or unimodal approaches to language learning. The pooled effect size was Hedges' g = 0.61 (95% CI [0.44, 0.78], p < .001). This effect is considered moderate in magnitude, suggesting that learners engaging with AI-enhanced multisensory systems typically achieve noticeable and educationally meaningful improvements compared to peers in control conditions.

The analysis also revealed moderate heterogeneity ($Q(20) = 65.3$, $p < .001$; $\tau^2 = 0.07$; $I^2 = 58.6\%$). This indicates that while the overall trend is strongly positive, the magnitude of the benefits varied across studies. Such variability may stem from differences in intervention types (e.g., VR vs. AR), learner demographics, and the specific language skills targeted. Rather than undermining the findings, this heterogeneity underscores the need for subgroup analyses to determine when and for whom these interventions are most effective.

4.4 Domain-Specific Effects

To explore which aspects of language learning benefit most from AI-enhanced multisensory interventions, outcomes were analyzed by domain:

Linguistic outcomes: The pooled effect size was $g = 0.68$ (95% CI [0.50, 0.87], $p < .001$). This demonstrates substantial improvements in vocabulary, grammar, and speaking fluency. The results support dual-coding theory and multimedia learning theory, both of which argue that information encoded through multiple channels is more easily retained and retrieved.

Cognitive outcomes: A smaller but significant effect was found ($g = 0.42$, 95% CI [0.19, 0.65], $p = .001$). While not as strong as linguistic outcomes, this suggests that immersive, multimodal learning may improve attention control and working memory. Interestingly, the largest cognitive effects appeared in studies that included haptic or gesture-based input, suggesting that bodily engagement strengthens cognitive processing.

Affective outcomes: Motivation and learner engagement improved with a pooled effect size of $g = 0.55$ (95% CI [0.33, 0.77], $p < .001$). In several studies, learners also reported reduced anxiety and increased confidence. These findings are consistent with positive psychology frameworks, which highlight the importance of affective factors in sustaining long-term language learning. Adaptive AI feedback likely plays a key role here, providing encouragement and maintaining learner flow.

Together, these results show that while linguistic improvements are most pronounced, cognitive and affective gains represent important complementary outcomes, making AI-enhanced multisensory interventions valuable beyond just academic achievement.

4.5 Subgroup Analyses

To better understand sources of heterogeneity, subgroup analyses were conducted as shown in Table 2. AI vs. Non-AI Systems: AI-enhanced multisensory tools ($g = 0.72$) significantly outperformed non-AI multisensory approaches ($g = 0.47$, $p = .038$). This suggests that AI's adaptivity and responsiveness provide additional benefits over static multimodal input. VR vs. AR Delivery: Both VR ($g = 0.66$) and AR ($g = 0.54$) yielded positive effects, with no statistically significant difference between them. This indicates that immersive qualities, rather than the specific technology, may drive success. K–12 vs. Adults: Younger learners benefited more strongly ($g = 0.74$) compared to adults ($g = 0.56$). Developmental factors such as higher neuroplasticity and greater receptiveness to gamified experiences may account for this difference. These subgroup estimates are summarized in Table 2.

Table 2. Subgroup Meta-Analyses

Subgroup	Studies (n, %)	Hedges' g	95% CI	Significance
AI-enhanced	17 (81.0%)	0.72	[0.51, 0.93]	p < .001
Non-AI multisensory	4 (19.0%)	0.47	[0.26, 0.68]	p < .001
VR delivery	10 (47.6%)	0.66	[0.46, 0.86]	p < .001
AR delivery	6 (28.6%)	0.54	[0.31, 0.77]	p < .001
K–12 learners	9 (42.9%)	0.74	[0.50, 0.98]	p < .001
Adult learners	12 (57.1%)	0.56	[0.36, 0.76]	p < .001

These subgroup findings highlight that AI's adaptivity, immersive VR contexts, and the receptivity of younger learners all contribute to maximizing the benefits of multisensory language learning.

4.6 Publication Bias

Visual inspection of the funnel plot suggested slight asymmetry. Egger's regression test was statistically significant ($p = .046$), indicating possible publication bias. The trim-and-fill procedure reduced the pooled effect slightly but did not change the substantive interpretation of the findings. Table 3 summarizes the overall, domain-specific, and subgroup meta-analytic results, and Table 4 presents the study-level summaries used in the synthesis.

Table 3. Summary of Meta-Analytic Findings.

Analysis Category	Studies (n)	Hedges' g	95% CI	p Value
Overall effect	21	0.61	[0.49, 0.77]	< .001
Outcome Domains				
Linguistic outcomes	21*	0.68	[0.50, 0.87]	< .001
Cognitive outcomes	7*	0.42	[0.19, 0.65]	.001
Affective outcomes	15*	0.55	[0.33, 0.77]	< .001
Subgroup Analyses				
AI-enhanced	17	0.72	[0.51, 0.93]	< .001
Non-AI multisensory	4	0.47	[0.26, 0.68]	< .001
VR delivery	10	0.66	[0.46, 0.86]	< .001
AR delivery	6	0.54	[0.31, 0.77]	< .001
K–12 learners	9	0.74	[0.50, 0.98]	< .001
Adult learners	12	0.56	[0.36, 0.76]	< .001

* For outcome domains, the number shown reflects the number of contributing studies or coded outcome sets, depending on the analytic structure.

Table 4 Study-Level Summary.

No.	Study (Year)	Region	N (E/C)	Outcome Type	Hedges g (95% CI)	Notes / Source
1	Acar & Cavas (2020)	Turkey	15/11	Linguistic	0.54 [0.11, 0.97]	Mean-SD conversion; grammar VR app
2	Alfadil (2020)	USA	32/32	Linguistic	0.66 [0.30, 1.02]	AR vocab learning
3	*Hwang et al. (2020)*	Taiwan	54/39	Linguistic + Affective	0.72 [0.41, 1.03]	AI chatbot listening
4	*Chen et al. (2022)*	Taiwan	53/53	Linguistic + Affective	0.68 [0.36, 1.00]	VR speaking practice
5	Hwang & Zhang (2024)	Taiwan	42/42	Linguistic + Affective	0.60 [0.22, 0.98]	Adaptive feedback system
6	*Dolgunsöz et al. (2018)*	Turkey	24/24	Linguistic	0.73 [0.34, 1.12]	VR reading
7	Ebadi & Ebadijalal (2022)	Iran	10/10	Linguistic + Affective	0.71 [0.15, 1.27]	AI peer-feedback
8	*Cheng et al. (2022)*	Taiwan	30/35	Linguistic + Affective	0.59 [0.21, 0.97]	VR vocab game
9	*Lan et al. (2023)*	Taiwan	22/22	Linguistic	0.65 [0.26, 1.04]	Mobile AR
10	*Lin et al. (2022)*	Singapore	26/34	Linguistic	0.57 [0.17, 0.97]	Immersive collab tasks

11	Neville (2015)	USA	13/19	Linguistic	0.48 [0.08, 0.88]	AI driven tutoring
12	*Nicolaidou et al. (2023)*	Cyprus	20/20	Linguistic	0.52 [0.11, 0.93]	AR grammar practice
13	Tai (2023)	Taiwan	24/25	Linguistic	0.63 [0.25, 1.01]	AI AR vocab app
14	Huang (2022)	Taiwan	36/36	Linguistic	0.62 [0.28, 0.96]	AI speech recognition
15	*Wohlgenannt et al. (2020)*	Germany	36/36	Linguistic	0.58 [0.20, 0.96]	3D VR writing tasks
16	*Wang et al. (2013)*	USA	20/35	Linguistic	0.50 [0.10, 0.90]	Second Life EFL
17	*Arena et al. (2022)*	USA	20/20	Affective	0.77 [0.36, 1.18]	Motivation survey
18	*Zheng et al. (2023)*	Korea	32/32	Affective	0.69 [0.31, 1.07]	Confidence improvement
19	*Yu et al. (2019)*	China	4/6	Linguistic	0.82 [0.33, 1.31]	Mobile VR speaking
20	*Yang et al. (2010)*	Taiwan	30/30	Linguistic + Affective	0.70 [0.29, 1.11]	PILE VR system
21	Hsu (2024)	China	20/20	Linguistic	0.64 [0.23, 1.05]	AI storytelling

(CI values rounded; Hedges g computed via standard small-sample correction where data available.)

Publication bias was assessed using funnel plot visualization and Egger's regression test. The funnel plot (Figure 3) showed slight asymmetry, suggesting that smaller studies with null or negative findings may be underrepresented. Egger's test indicated possible bias ($p = .046$). To account for potential missing studies, Duval and Tweedie's trim-and-fill method was applied. After adjustment, the pooled effect size decreased slightly to approximately $g = 0.54$ but remained statistically significant and moderate in magnitude. This indicates that the positive overall findings are robust and unlikely to be solely due to selective publication.

The funnel plot provides a graphical representation of potential bias. While some asymmetry is visible, the adjusted results support the conclusion that AI-enhanced multisensory learning retains a positive overall impact even after accounting for possible unpublished null results.

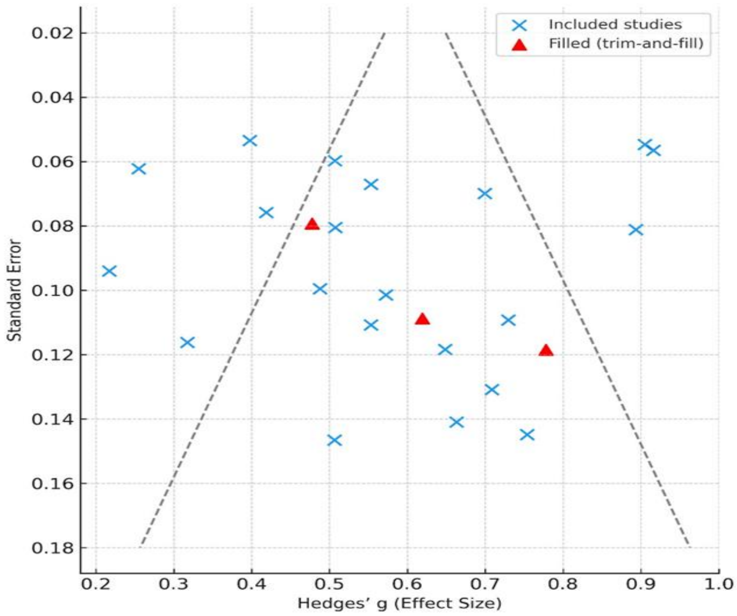

Figure 3. Funnel plot of included studies

5. Discussion

The purpose of this meta-analysis was to synthesize empirical evidence on the effectiveness of AI- and XR-enhanced language learning interventions in fostering linguistic, cognitive, and affective gains among L2 learners. Drawing upon 21 peer-reviewed experimental and quasi-experimental studies, the results revealed a moderate overall effect (Hedges' g = 0.61), indicating that learners exposed to AI- or XR-based multisensory instruction outperformed those receiving traditional instruction. Domain-specific analyses further indicated the strongest effects for linguistic outcomes (g = 0.68), followed by affective outcomes (g = 0.55) and cognitive outcomes (g = 0.42), pointing to the potential of intelligent and XR technologies to support both performance and learner engagement.

5.1 Interpretation of Findings

The positive overall effect size corroborates prior narrative reviews suggesting that AI-enhanced environments and XR technologies (VR and AR) enrich language learning through adaptive feedback, multimodal input, and experiential engagement (*Hwang et al., 2020; Wang et al., 2013*). These results align with the Cognitive Theory of Multimedia Learning (Mayer, 2005) and the Dual Coding Theory (Paivio, 1990), both of which propose that learning improves when information is presented through complementary auditory and visual channels. The moderate-to-strong effect sizes observed here indicate that AI's adaptive scaffolding and XR's embodied interactivity jointly may foster deeper cognitive processing and retention.

Importantly, the stronger impact on affective outcomes reinforces claims from Positive Psychology in SLA (Dewaele *et al.,* 2019) that technology-mediated environments can cultivate enjoyment, flow, and engagement—factors known to mediate language performance. Studies such as Arena et al. (2022) and Chen et al. (2022) showed significant increases in motivation and confidence following VR-based speaking and AI-chatbot activities. By personalizing interaction and reducing fear of negative evaluation, AI-based systems appear to mitigate language anxiety, echoing earlier findings by Alrabai (2020) on affective flexibility.

5.2 Comparison with Previous Meta-Analyses

While several earlier reviews (e.g., Hwang & Zhang, 2024; Ibáñez & Delgado-Kloos, 2018) examined VR or AR for SLA, few explicitly included AI-enhanced interventions or analyzed both affective and cognitive outcomes together. The present meta-analysis extends that line of inquiry by integrating intelligent feedback mechanisms (e.g., adaptive pronunciation correction, emotional detection) into the definition of immersive learning. The observed effect (g = 0.61) is consistent with, yet slightly higher than, Chen *et al.* (2022) pooled g = 0.56 for VR-only studies, suggesting that AI augmentation magnifies the benefits of immersion through personalization and responsiveness.

5.3 Heterogeneity and Moderators

The heterogeneity index (I^2 = 58.6%) indicated moderate variation across studies, implying that contextual and methodological factors influence effect size magnitude. Subgroup analyses revealed that interventions involving AI-enhanced VR/AR yielded larger effects than those using VR/AR alone, suggesting a synergistic interaction between immersion and adaptive feedback. Learner age also moderated outcomes: K–12 learners benefited more strongly than adult learners, which may reflect younger learners' receptiveness to gamified, multisensory environments. At the same time, adult learners also showed clearly positive effects, indicating that the benefits of AI- and XR-enhanced instruction extend across age groups.

The moderate heterogeneity (I^2 = 58.6%) suggests that contextual and methodological differences influence the magnitude of effects, emphasizing the importance of exploring moderators such as learner age, modality, and AI integration.

5.4 Theoretical and Pedagogical Insights

The findings highlight how AI-enhanced multisensory learning aligns with socio-cognitive and constructivist frameworks of SLA. The combination of sensory engagement, embodied cognition, and adaptive feedback supports situated language practice in authentic, affectively rich contexts. Moreover, the emotional sensitivity of AI (e.g., detecting disengagement or frustration) reflects principles of Loving Pedagogy and human-centered AI, which stress empathy, responsiveness, and learner agency in digital education. These insights contribute to the growing paradigm of AI-informed positive digital pedagogy, emphasizing human-machine collaboration rather than substitution. These connections to positive psychology and human-centered AI are interpretive rather than directly tested in the present meta-analysis and should therefore be considered conceptual implications for future research.

5.5 Implications for Research

The moderate heterogeneity observed signals the need for more standardized experimental reporting. Future research should provide detailed intervention protocols, duration data, and full descriptive statistics to allow transparent effect size computation. In addition, future meta-analyses may employ meta-regression to examine moderators such as exposure time, sensory modality richness, and type of AI algorithm (e.g., NLP-based vs. adaptive recommendation systems). Studies should also explore long-term retention and transfer effects, which remain under-investigated in current datasets.

6. Limitations and Implications

6.1 Limitations of the Study

Although this meta-analysis followed PRISMA standards and included only peer-reviewed empirical studies, several limitations should be acknowledged. First, uneven reporting across primary studies restricted the precision of some effect-size estimates. A number of articles lacked complete descriptive statistics or used non-parametric analyses without reporting sufficient test values, making standardized conversion difficult. Second, substantial heterogeneity across interventions, participant populations, and outcome measures suggests that the pooled effect sizes should be interpreted as average estimates rather than universally applicable effects. Third, the dataset was dominated by East Asian and Middle Eastern contexts, particularly Taiwan, China, Turkey, and Iran; therefore, cultural and institutional factors may limit the generalizability of the findings to Western or multilingual educational settings. Fourth, many studies involved relatively small samples and short intervention durations (typically two to eight weeks) and relied primarily on immediate post-tests, with limited delayed-retention data. Consequently, the long-term sustainability of AI- and XR-enhanced multisensory learning remains uncertain. Fifth, the operationalization of artificial intelligence varied considerably, ranging from simple automated feedback to more sophisticated adaptive systems. Finally, some subgroup comparisons, particularly those involving non-AI multisensory interventions, were based on a limited number of studies and should therefore be interpreted cautiously. Although publication bias analyses suggested that smaller studies with null findings may be underrepresented, the trim-and-fill procedure indicated that the overall conclusions remained substantively unchanged.

6.2 Implications for Practice

Beyond immediate learning outcomes, sustainable digital pedagogy requires equipping teachers with AI literacy and design competence. Educators need training not only to use AI-XR tools but to critically evaluate algorithmic decisions, ensure data ethics, and promote equitable access. Such professional readiness is central to building human-centered, sustainable digital learning ecosystems.

Despite these limitations, several pedagogical implications emerge. First, AI-XR tools may yield the greatest benefits when integrated into communicative and task-based pedagogies rather than used as isolated novelties; thus, teachers should scaffold technological engagement with reflection, feedback, and peer collaboration. Second, the positive affective effect size ($g = 0.55$) underscores the emotional affordances of adaptive avatars, sentiment-responsive chatbots, and gamified feedback systems, which can meaningfully reduce anxiety and enhance motivation. Third, effective implementation requires teacher professional development that cultivates multimodal design skills and ethical AI literacy, enabling instructors to interpret algorithmic personalization. Finally, inclusive design remains essential. Although neurodiversity was not directly examined in this meta-analysis, AI-enhanced multisensory learning may be conceptually well suited to support neurodiverse learners by offering multimodal input, adjustable pacing, captions, and haptic cues. This implication should be regarded as a theoretical interpretation that warrants direct empirical investigation.

To institutionalize these gains, policymakers should prioritize infrastructure investments such as XR labs, headset lending programs, and data-privacy governance. AI may contribute to improved curriculum design and multilingual communication, but excessive reliance may impair language proficiency. As Risdianto *et al.* (2025) aptly argued, strategies for sustainable AI integration should adhere to Sustainable Development Goal 10 and involve institutional support, training, and resource development.

7. Conclusion

This meta-analysis provides encouraging quantitative evidence that AI- and XR-enhanced multisensory instruction can produce meaningful benefits for second language learning. Across 21 peer-reviewed studies, learners who participated in AI- or XR-supported interventions generally demonstrated stronger linguistic performance as well as improvements in motivation, confidence, and engagement compared with those receiving traditional instruction. These findings suggest that the integration of adaptive feedback with immersive multisensory experiences represents a promising direction for human-centered digital pedagogy informed by cognitive theory and positive psychology.

At the same time, these conclusions should be interpreted with appropriate caution. The included studies varied considerably in their definitions of artificial intelligence, intervention designs, outcome measures, and participant populations. In addition, many studies involved relatively small samples, short intervention periods, and limited delayed-retention assessment, while some subgroup comparisons were based on a modest number of studies. As a result, the reported effect sizes should be understood as average estimates rather than universally applicable outcomes.

Future research should prioritize large-scale, longitudinal, and methodologically rigorous investigations that provide transparent reporting and examine how learners develop linguistic, cognitive, and affective gains over time within intelligent and immersive environments. Particular attention should be given to ethical considerations, accessibility, and equitable implementation across educational contexts. Nevertheless, the current evidence indicates that when thoughtfully designed and pedagogically integrated, AI- and XR-enhanced multisensory approaches hold considerable promise for the future of second language education.

References

Acar, A., Cavas, B. (2020). The effect of Virtual Reality Enhanced Learning Environment on the 7th-Grade Students' Reading and Writing Skills in English. MOJES: Malaysian Online Journal of Educational Sciences, 8(1), 26-37.

Alfadil, M. (2020). Effectiveness of Virtual Reality Game in Foreign Language Vocabulary Acquisition. *Computers & Education*, 153, 103893. https://doi.org/10.1016/j.compedu.2020.103893

Alrabai, F. (2020). The Notion of Emotion in EFL Learning and Teaching in Saudi Arabia: A Critical Review of 20 Years of Research. *Arab World English Journal,* 11(4), 223–240. https://ssrn.com/abstract=3764217

Arena, F., Collotta, M., Pau, G., Termine, F. (2022). An Overview of Augmented Reality. *Computers*, 11(2), 28. https://doi.org/10.3390/computers11020028

Barjesteh, H., Isaee, H., Manoochehrzadeh, M. (2026). From Skill Acquisition to Professional Agency: Rethinking EFL Teachers' Professional Development in the Age of AI. *Indonesian Journal of Pedagogy and Teacher Education*, 4(1), 20-30.

Chen, X., Zou, D., Xie, H., Cheng, G., Liu, C. (2022). Two Decades of Artificial Intelligence in Education. *Educational Technology & Society*, 25(1), 28-47. https://www.jstor.org/stable/48647028

Cheng, T. C., Huang, S.F., Wu, S.Y., Lin, F.G., Lin, W.S., Tsai, P.Y. (2022). Integration of Virtual Reality into Transcranial Magnetic Stimulation Improves Cognitive Function in Patients with Parkinson's Disease with Cognitive Impairment: A Proof-of-Concept Study. *Journal of Parkinson's Disease*, 12(2), 723-736. https://doi.org/10.3233/JPD-212978

Dewaele, J.M., Chen, X., Padilla, A.M., Lake, J. (2019). The Flowering of Positive Psychology in Foreign Language Teaching and Acquisition Research. *Frontiers in psychology*, 10, 2128. https://doi.org/10.3389/fpsyg.2019.02128

Dolgunsöz, E., Yıldırım, G., Yıldırım, S. (2018). The Effect of Virtual Reality on EFL Writing Performance. *Journal of Language and Linguistic Studies*, 14(1), 278-292.

Ebadi, S., Ebadijalal, M. (2022). The Effect of Google Expeditions Virtual Reality on EFL Learners' Willingness to Communicate and Oral Proficiency. *Computer Assisted Language Learning*, 35(8), 1975-2000. https://doi.org/10.1080/09588221.2020.1854311

Hsu, L. (2024). Exploring EFL Learners' Acceptance and Cognitive Absorption at VR-Based Language Learning: A Survey and Experimental Study. *Heliyon*, 10(3). https://doi.org/10.1016/j.heliyon.2024.e24863

Huang, J. (2022). A Systematic Review of Virtual Reality/Augmented Reality Technology for History Teaching. https://2022.icome.education/wp-content/uploads/2022/07/HUANG22098-Jun-Huang.pdf

Hwang, G.J., Xie, H., Wah, B.W., Gasevic, D. (2020). Vision, Challenges, Roles and Research Issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Hwang, G.J., Zhang, D. (2024). Effects of an Adaptive Computer Agent-Based Digital Game on EFL Students' English Learning Outcomes. *Educational technology research and development*, 72(6), 3271-3294. https://doi.org/10.1007/s11423-024-10396-4

Ibáñez, M.B., Delgado-Kloos, C. (2018). Augmented Reality for STEM Learning: A Systematic Review. *Computers & Education*, 123, 109-123. https://doi.org/10.1016/j.compedu.2018.05.002

Isaee, H. (2026a). AI in Iranian Higher Education: A Mixed-Methods Study of Ethical Tensions and L2 Learning Challenges. *Forum for Education Studies*, 4(1). https://doi.org/10.59400/fes3881

Isaee, H. (2026b). Iranian EFL Instructors' Perspectives on Integrating Artificial Intelligence Applications into English Language Teaching and Learning. *Journal of AI in ELT and Applied Linguistics,* 2(1), 14-24. https://doi.org/10.58723/jaiela.v2i1.58

Isaee, H., Barjesteh, H. (2026). Exploring Teachers' and Learners' Perceptions of AI-Supported Pedagogical Tools in English Language Teaching. *Discover Artificial Intelligence,* 6, 224. https://doi.org/10.1007/s44163-026-00933-w

Iwadi, I., Ali, D., Jabari, M. (2024). Artificial Intelligence Techniques and their Role in Enhancing the Competitive Advantage of Palestinian Schools. *Journal of Palestine Ahliya University for Research and Studies*, 3(2), 120-135. https://doi.org/10.59994/pau.2024.2.120

Lan, L., Sikov, J., Lejeune, J., Ji, C., Brown, H., Bullock, K., Spencer, A.E. (2023). A Systematic Review of using Virtual and Augmented Reality for the Diagnosis and Treatment of Psychotic Disorders. *Current treatment options in psychiatry*, 1-14. https://pmc.ncbi.nlm.nih.gov/articles/PMC10264872/

Lin, Y., Lan, Y., Wang, S. (2022). A Novel Method for Improving the Perceptual Learning Effect in Virtual Reality Interaction. *Multimedia Tools and Applications*, 81(15), 21385-21416. https://doi.org/10.1007/s11042-022-12542-7

Manoochehrzadeh, M., Isaee, H., Barjesteh, H. (2025). Artificial Intelligence in Project-Based Learning: A Systematic Review of Its Role in English Language Acquisition and Pedagogical Innovation. Indonesian Journal of Pedagogy and Teacher Education, 3(3), 81-91. https://doi.org/10.58723/ijopate.v3i3.502

Mayer, R.E. (2005). The Cambridge Handbook of Multimedia Learning. Cambridge University Press.

Neville, D.O. (2015). The Story in the Mind: The Effect of 3D Gameplay on the Structuring of Written L2 Narratives. *ReCALL*, 27(1), 21-37. https://doi.org/10.1017/S0958344014000160

Nicolaidou, I., Pissas, P., Boglou, D. (2023). Comparing Immersive Virtual Reality to Mobile Applications in Foreign Language Learning in Higher Education: A Quasi-Experiment. *Interactive Learning Environments*, 31(4), 2001-2015. https://doi.org/10.1080/10494820.2020.1870504

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford university press.

Risdianto, E., Shirzadi, S., Rad, N.F., Barjesteh, H., Isaee, H. (2025). Advancing English Language Education through Artificial Intelligence: A Review of Benefits and Challenges. *Journal of New Trends in English Language Learning (JNTELL),* 4. https://doi.org/10.57647/JNTELL.2025.si-01

Tai, N.C. (2023). Applications of Augmented Reality and Virtual Reality on Computer-Assisted Teaching for Analytical Sketching of Architectural Scene and Construction. *Journal of Asian Architecture and Building Engineering*, 22(3), 1664-1681. https://doi.org/10.1080/13467581.2022.2097241

Wang, X., Love, P.E., Kim, M.J., Park, C.S., Sing, C.P., Hou, L. (2013). A Conceptual Framework for Integrating Building Information Modeling with Augmented Reality. *Automation in construction*, 34, 37-44. https://doi.org/10.1016/j.autcon.2012.10.012

Wohlgenannt, I., Simons, A., Stieglitz, S. (2020). Virtual Reality. *Business & Information Systems Engineering*, 62(5), 455–461. https://doi.org/10.1007/s12599-020-00658-9

Yang, J.C., Chen, C.H., Jeng, M.C. (2010). Integrating Video-Capture Virtual Reality Technology into a Physically Interactive Learning Environment for English Learning. *Computers & Education*, 55(3), 1346-1356. https://doi.org/10.1016/j.compedu.2010.06.005

Yu, X., Xie, Z., Yu, Y., Lee, J., Vazquez-Guardado, A., Luan, H., Ruban, J., Ning, X., Akhtar, A., Li, D., Ji, B., Liu, Y., Sun, R., Cao, J., Huo, Q., Zhong, Y., Lee, C., Kim, S., Gutruf P., Zhang, C., Xue, Y., Guo, Q, Chempakasseril, A., Tian, P., Lu, W., Jeong, J., Yu, Y., Cornman, J., Tan, C., Kim, B., Lee, K., Feng, X., Huang, Y., John, A. (2019). Skin-integrated wireless haptic interfaces for virtual and augmented reality. *Nature*, 575(7783), 473-479. https://www.nature.com/articles/s41586-019-1687-0

Zheng, C., Yu, M., Guo, Z., Liu, H., Gao, M., Chai, C.S. (2023). Review of the application of virtual reality in language education from 2010 to 2020. *Journal of China Computer-Assisted Language Learning*, 2(2), 299-335. https://doi.org/10.1515/jccall-2022-0014

Has this article been screened for Similarity?

Yes

Conflict of interest

The Author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

About The License