12th International Conference of the Immersive Learning Research Network
Selected Academic papers for the iLRN Proceedings

DOI: https://doi.org/10.56198/S7USIPGD9

Virtual Reality in Language Education: Evidence from a Systematic Review

Matthias Conrad[1] [0000-0003-1063-2231] and Mark Schaermann[1]

[1]University of Konstanz, Konstanz, Germany
matthias.conrad@uni-konstanz.de

Abstract. Immersive Virtual Reality (VR) enables the simulation of authentic scenarios that can support the acquisition of foreign language skills. This systematic review synthesizes empirical evidence on the use of VR in language education and adopts the Common European Framework of Reference for Languages (CEFR) as an organizing framework for structuring language skills. Following PRISMA guidelines, 25 peer reviewed studies published between 2010 and 2022 were identified and analyzed.

Overall, the findings indicate that VR can support language learning outcomes, although the magnitude and consistency of effects vary across skill domains. More consistent positive effects are observed for speaking and listening, where immersive and multimodal environments appear to facilitate comprehension, fluency, and learner confidence. Improvements in reading and writing are reported less consistently and seem to depend on structured instructional design and pedagogical scaffolding. Evidence regarding vocabulary learning is mixed and strongly influenced by task characteristics, exposure time, and comparison conditions.

Taken together, the results largely converge with findings reported in other systematic reviews on VR in language learning. However, by explicitly situating effects along CEFR skill domains, the present review offers a more differentiated and pedagogically grounded perspective on where and under which conditions immersive VR appears particularly supportive.

Keywords: VR, Education, Language, Training, Communication, Skills.

1 Introduction

The use of immersive virtual reality (VR) has become well established in diverse educational and training contexts. Its didactic application in foreign language teaching is widely regarded as particularly promising for fostering communicative skills. In addition, immersive environments can enhance learners' self-efficacy and reduce foreign language anxiety by providing a safe space for practice and experimentation [11, 12, 32]. This potential is closely linked to effective instructional design, which structures content and supports comprehension, processing, and retention. Integrating VR into language learning can not only improve learning outcomes but can also increase learners' motivation and positive perceptions of the learning environment [16].

Although research on VR based language learning highlights the general promise of immersive technologies, it remains unclear which specific areas of language acquisition benefit most from their use. Addressing this question requires a structured and transparent framework that allows for the systematic organization and interpretation of empirical findings. Against this background, the present systematic review synthesizes existing evidence and critically examines the effectiveness of immersive VR across different dimensions of language learning, identifying both pedagogical potentials and limitations.

Accordingly, this study adopts the *Common European Framework of Reference for Languages* (CEFR) [7] and the CEFR *Companion Volume* [8] as its conceptual foundation. While the CEFR provides the overarching proficiency framework, the *Companion Volume* extends this perspective through descriptors for interaction, mediation, and digitally mediated communication. In the following, the CEFR is used as an umbrella concept referring jointly to the core framework and its extended descriptor framework, ensuring terminological consistency throughout the manuscript. It distinguishes four interrelated skills: *reading comprehension*, *listening comprehension*, *writing*, and *speaking*. In addition, the framework defines six proficiency levels ranging from A1 to C2, providing a differentiated structure for describing linguistic competence. Building on this framework, the

present systematic review adopts the CEFR structure to categorize and examine the effectiveness of immersive VR interventions across language skills and proficiency levels. By doing so, the study contributes to existing research not merely by aggregating prior reviews, but by critically synthesizing and extending their findings for the description, evaluation, and classification of language proficiency. In contrast to other reviews that primarily emphasize overall effectiveness, motivation, or technological aspects [9, 11, 16, 32], the present analysis places greater emphasis on learning goals, target groups, and educational contexts.

2 Research Questions

This study aims to identify evidence based findings regarding the efficacy of immersive VR applications for foreign language learning. Grounded in the fundamental skill domains outlined in the *Common European Framework of Reference for Languages* (CEFR), the core research question examines the effects of VR in foreign language teaching on the development of *reading comprehension*, *listening comprehension*, *writing*, *speaking*, and *vocabulary*.

By systematically addressing this question, the study seeks to generate comprehensive insights into the effectiveness of VR based language instruction in achieving specific learning objectives. These insights are intended to advance both technology oriented educational research and pedagogical practice in immersive language learning contexts. Ultimately, the findings may contribute to a more targeted and pedagogically informed use of VR as a digital teaching and learning medium.

3 Method

To address the research questions, a systematic review of VR based language learning was conducted in 2022 following PRISMA guidelines [23, 28]. A sensitive search strategy was employed to maximize coverage and minimize the risk of overlooking relevant studies. To this end, the research topic was decomposed into four core search components: learning effects, VR, head mounted display technologies, and foreign language learning. Keywords were derived from these components in a transparent and reproducible manner. Relevant terms and synonyms were identified through scoping searches and prior literature and were systematically combined using Boolean operators to construct database specific search strings (Table 1). This approach ensured high sensitivity while maintaining adequate precision in identifying empirical studies at the intersection of immersive VR and language learning.

Table 1. Search components and keywords.

Search components	Keywords		
Effect / Impact	- effect - impact - cognition - competence - educational achievement - educational effectiveness - educational outcome - educational success	- expertise - know - know-how - knowledge - learn - learning achievement - learning effectiveness - learning outcome - learning performance learning result - learning success	- result - outcome in learning - outcome of learning - performance - skill - solve - study achievement - study effectiveness - study outcome - study result - study success
Head-Mounted-Display (HMD)	- cardboard - goggles - head mounted display - head mounted helmet	- headset - headgear-driven - head-mounted display based - headset-driven	- HMD - HTC Vive - immersive - Oculus - Samsung Gear

Foreign Language Learning	- foreign language learning - foreign language education - L2 - L2 learning - second language acquisition - second language learning - L2 education	- L2 acquisition - reading - writing - pronunciation - mediation language mediation - listening - communication - vocabulary - grammar - intonation	- oral - reading comprehension reading skills - writing skills - listening comprehension - communication skills - oral skills/ oral proficiency
Virtual Reality (VR)	- virtual reality - VR - virtual environment - VE - IVE - IVLE - VLE - virtual world - virtual simulation	- virtual learning environment - VR-technology - VR technology - Immersive technology - 3D simulation technology	- interactive simulation technology - spatial computing technology - digital reality technology

3.1 Selection of Databases and Search Strategy

To ensure a comprehensive and methodologically sound search, multiple academic databases were systematically consulted, as individual databases differ in scope and coverage. Database selection was guided by the interdisciplinary nature of the research topic, which is situated at the intersection of educational research, psychology, and foreign language didactics. Accordingly, databases with extensive coverage of peer reviewed research in these domains were selected. The search included *APA PsycArticles* and *APA PsycInfo* as key sources for empirical and theoretical research in psychology and the learning sciences, *ERIC* as a core database for educational research, and *Teacher Reference Center* as a complementary source for practice oriented and pedagogically focused studies in foreign language education. The combined use of these databases ensured broad disciplinary coverage and enhanced the transparency and rigor of the search process. All databases were accessed via the EBSCOhost platform, which enables integrated cross database searching and the consistent application of search strings and filters. This approach facilitated an efficient, transparent, and reproducible literature search.

The screening and selection of studies were conducted by a single reviewer, who was responsible for title and abstract screening, full text assessment, and final inclusion decisions based on predefined criteria. While this approach precluded the calculation of interrater agreement, it ensured consistency in the application of selection criteria across all stages of the review process.

3.2 Inclusion and Exclusion Criteria

In terms of the underlying VR technology, only studies employing a head mounted display or an HMD like device were included in order to ensure a sufficient level of immersion and to exclude non immersive desktop based VR applications. Only peer reviewed journal articles published from 2010 onward were considered, as earlier studies often refer to outdated technologies. Dissertations, book chapters, reports, and conference presentations were excluded. Eligible articles were required to report an empirical study and to contain the relevant search terms in the title. These inclusion and exclusion criteria were applied to ensure a high scientific standard and methodological robustness.

Following the screening process, 25 studies met all criteria. An overview of the selection procedure is provided in the PRISMA flowchart (Figure 1). The identified studies were then systematically analyzed according to the CEFR skill domains. In addition to reading, listening, writing, and speaking, vocabulary was included as an additional analytical category due to its prominence in the reviewed studies.

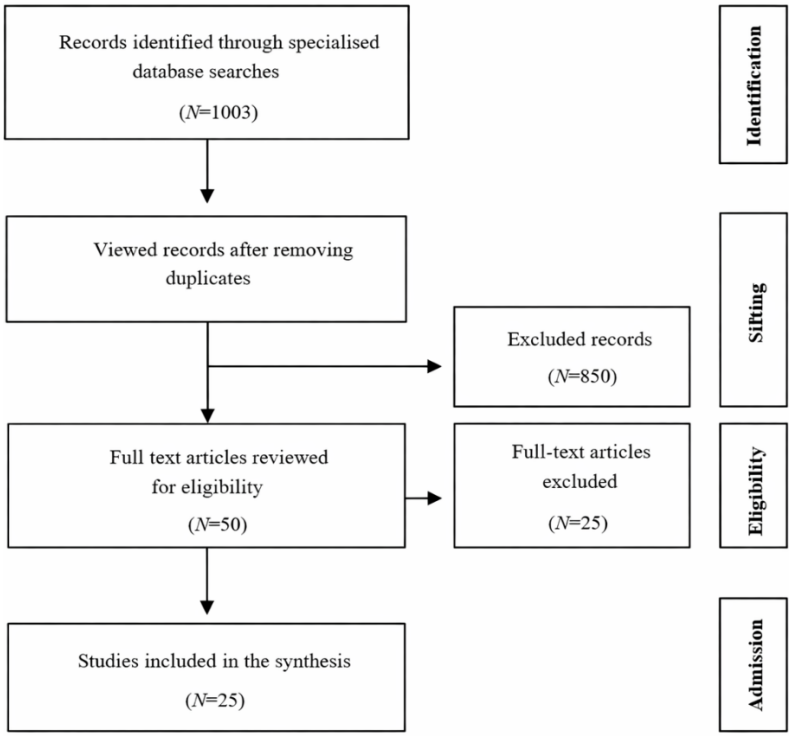

Fig. 1. Flowchart of the PRISMA selection process (original graphic).

4 Results

Table 2 provides an overview of the 25 identified studies. Sample sizes ranged from small exploratory groups with approximately a dozen participants to larger cohorts comprising more than one hundred learners. Most studies were conducted in Asia, particularly in Taiwan, China, and Turkey, with a smaller number originating from Europe and North America. Participants' ages varied considerably, spanning primary, secondary, and tertiary education, with the majority of studies focusing on learners in secondary and tertiary contexts. The investigated target languages were predominantly English and Chinese, while a small number of studies examined learning in other languages. In terms of research design, quasi experimental approaches were most common, followed by mixed method designs.

Table 2. Characteristics of the identified studies.

	N	Ø Age	Country	Target Language	Level of Education	Area of expertise	Research design
Acar & Cavas, 2020	26	12	Turkey	English	Secondary	Reading & Writing	QED
Chen & Yuan, 2023	30	20.1	China	Chinese	Tertiary	Vocabulary	MM
Chen et al, 2021	84	n.s	Taiwan	English	Tertiary	Vocabulary	MM
Cheng et al, 2017	68	n.s.	USA	Japanese	Tertiary	Vocabulary	QED
Chen et al, 2022	59	10-11	China	Chinese	Primary	Writing	QED
Chien et al., 2020	69	16	Taiwan	English	Secondary	Communication	QED
Dolgunsöz et al., 2018	24	18-19	Turkey	English	Tertiary	Writing	MM
Hartfill et al., 2020	29	25.8	Germany	Japanese	n.s.	Vocabulary	MM
Hsu et al., 2021	131	18-22	Taiwan	English	Tertiary	Reading & Vocabulary	QED
Huang et al., 2020	65	17	Taiwan	Chinese	Secondary	Writing	MM
Khodabandeh, 2020	80	18-24	Iran	English	Tertiary	Writing	QED

Lai & Chen, 2021	30	17-18	China	English	Secondary	Vocabulary	MM
Lee, 2019	54	20-34	South Korea	English	n.s.	Listening	MM
Legault et al., 2019	64	19.05	USA	Chinese	Tertiary	Vocabulary	QED
Liu et al., 2023	63	20	China	English	Tertiary	Reading	QED
Madini & Alshaikhi, 2017	20	23-25	Saudi-Arabia	English	n.s.	Vocabulary	QED
Nicolaidou et al., 2021	40	19.6	Greece	Italian	Tertiary	Vocabulary	QED
Papin & Kaplan-Rakowski, 2022	63	19.7	Canada	French	Tertiary	Vocabulary	QED
Repetto et al., 2021	104	n.s.	Italy	English	Secondary	Vocabulary	QED
Tai & Chen, 2021	49	13-15	Taiwan	English	Secondary	Listening	MM
Tai et al., 2022	49	14-15	Taiwan	English	Secondary	Vocabulary	MM
Wang et al., 2021	98	19	China	English	Tertiary	Reading	QED
Xie et al., 2021	12	n.s.	USA	Chinese	Tertiary	Communication	MM
Yang et al., 2020	72	n.s.	Taiwan	English	Secondary	Communication	MM
Yang et al., 2021	40	10	China	Chinese	Primary	Writing	MM

MM = Mixed-Method; QED = Quasi-experimental

The results reported in detail below focus on studies employing experimental and quasi experimental control group designs that allow systematic comparisons between immersive VR interventions and non immersive or traditional instructional formats. Accordingly, the reported effect sizes reflect relative differences between instructional conditions and should be interpreted as comparative indicators of the effectiveness of HMD based immersive VR rather than as absolute learning gains.

Effect sizes were reported using *Cohen's d* as a standardized measure of group differences or pre post changes. When sufficient statistical information was available, *Cohen's d* was calculated by dividing the mean difference by the pooled standard deviation, computed as the square root of the weighted average of group variances. Where necessary, alternative statistics were converted into *Cohen's d* using established transformation formulas. Following conventional guidelines, values between 0.20 and 0.50 were interpreted as *small* effects, values between 0.50 and 0.80 as *medium* effects, and values above 0.80 as *large* effects.

The methodological quality of the included studies was not systematically assessed using standardized appraisal tools. Consequently, potential sources of bias related to study design, sampling, measurement, or intervention implementation cannot be ruled out, and comparisons of effect sizes across studies should be interpreted with caution.

4.1 Reading Comprehension

Four studies examined the use of VR for developing learners' reading comprehension skills. Two of these studies are also discussed in the section on writing competence, as they addressed multiple literacy skills [1, 14].

Acar and Cavas [1] compared immersive VR based writing activities with traditional instructional approaches using textbooks and Smartboard exercises. The results indicate that learners in the VR condition achieved significantly higher post test scores compared to their pre test performance ($d = .79$), suggesting a beneficial effect of VR on academic achievement in reading and writing in English. In contrast, the control group did not show any measurable improvement between the pre test and post test. However, the study assessed reading and writing jointly without clearly distinguishing between the two skills, which limits the precision of the reported findings with regard to reading comprehension.

In addition to gains in writing performance, Hsu et al. [14] also reported improvements in reading comprehension. Comparing learners' post test results with their pre test performance, they found a statistically significant increase in reading comprehension with a medium effect size ($d = .69$), indicating that the use of VR contributed to measurable reading gains.

Wang et al. [36] investigated the effects of VR using two experimental conditions. The first experimental group used HMDs with visual scaffolding, while the second used HMDs without scaffolding. Both groups were compared to a control group receiving traditional instruction (TI). The results revealed that the group using VR with scaffolding performed significantly better than the group without scaffolding. Moreover, significant differences among the three groups were found for the dimensions of locating information and text

comprehension, but not for evaluation and reflection. In particular, learners using the scaffolded VR approach achieved higher scores in locating information (d = .80) compared to both the non scaffolded VR group and the TI group. These findings suggest that VR environments providing embedded visual support are more effective for fostering reading comprehension than VR or traditional settings without scaffolding.

Liu et al. [24] emphasized that many learners struggle to understand textual meaning due to a lack of authentic learning contexts. To address this issue, they introduced an *Article Structure Strategy based Spherical Video VR approach* (ASS SVVR). This approach integrates spherical video based VR, which provides a 360 degree immersive environment, with a reading strategy that focuses on identifying signal words, recognizing text structures, and organizing main ideas to enhance comprehension and recall. Their findings show that learners who used the ASS SVVR approach outperformed those who learned through a conventional SVVR approach, with a large effect size (d = .80).

Taken together, these studies suggest that VR can support the development of reading comprehension, particularly when integrated with structured instructional strategies or scaffolding that guide learners in navigating and processing textual information.

4.2 Listening Comprehension

Two studies investigated the impact of VR on learners' listening comprehension [21, 35]. Lee [21] examined learners' performance across three listening test formats, namely audio, video, and VR tasks. Each format was administered to two learner groups representing higher and lower proficiency levels. The results revealed that learners in the VR condition performed better than those in both the video and audio conditions, regardless of proficiency level. Differences between the VR and audio conditions were statistically significant and associated with very large effect sizes (d = 3.10 for higher achieving learners and d = 2.70 for lower achieving learners). No statistically significant differences were found between the VR and video conditions. These findings suggest that visual and contextual information, whether provided through standard video or immersive VR, enhances listening performance compared to purely auditory input.

Tai et al. [35] conducted a similar investigation comparing a VR learning environment with a video based condition in which learners viewed instructional videos on smartphones and subsequently completed a listening comprehension test. The study aimed to determine whether VR could facilitate the listening comprehension of adolescent learners of English as a foreign language. Quantitative results revealed significant differences between the two groups in both the immediate post test and the delayed post test. The VR group consistently outperformed the video group, indicating that immersive learning contributed to improved comprehension and retention of auditory material. The reported effect sizes were moderate (d = .60 for the post test and d = .70 for the delayed post test), demonstrating that VR supported learners in processing and remembering spoken input more effectively than conventional video instruction.

Overall, these studies indicate that VR environments can enhance listening comprehension by combining auditory information with rich visual and contextual cues. This multimodal integration appears to facilitate deeper cognitive processing, leading to improved understanding and longer retention of listening material.

4.3 Writing

In the domain of writing performance, the findings are similarly mixed but generally indicate modest benefits of VR. Several studies reported higher mean scores for learners in VR conditions compared to control groups, although effect sizes were predominantly small.

Two studies compared HMD based learning with non VR classroom materials. Acar and Cavas [1] examined the effects of immersive VR supported writing activities in comparison with conventional instruction based on textbooks and Smartboard exercises. Their findings indicate that learners in the VR condition demonstrated significantly higher post test performance relative to their pre test results (d = .79), suggesting positive effects of VR on English reading and writing achievement. In contrast, no comparable performance gains were observed in the control group. However, as the study did not clearly differentiate between reading and writing outcomes, the interpretability of the reported effects remains limited. Khodabandeh [18] reported similar results, with the VR group outperforming the control group in the post test with a large effect size (d = 1.70).

Yang et al. [39] reported slight improvements in text structure (d = .40) and linguistic expression (d = .30), as well as a statistically significant but weak effect for thematic coherence (d = .30). Similarly, Huang et al. [15] observed better results for the VR group in the dimensions of content (d = .30) and presentation (d = .80), with the latter indicating a strong effect. No significant differences were found for organization or vocabulary use. Dolgunsöz et al. [10] reported slightly better overall results for learners in the traditional 2D video condition, although the difference was small (d = .30).

In summary, the findings suggest that VR can positively influence writing performance, although observed effects are often small and vary depending on the specific writing dimension assessed and the study design.

4.4 Speaking

Three studies investigated the use of VR for improving learners' communicative competence. All studies employed different comparison conditions, which limits the generalizability of the findings.

Chien et al. [6] compared a VR based learning condition that incorporated a peer assessment component with a VR condition without peer assessment. Their findings revealed that learners in the VR plus peer assessment condition achieved higher overall performance than those in the VR only condition (d = .40). More specifically, the experimental group performed better in fluency (d = .50), comprehension (d = .60), and linguistic maturity (d = .50), while no statistically significant differences were found for accuracy, intelligibility, or pronunciation. These results suggest that combining immersive environments with reflective peer feedback can positively influence certain dimensions of communicative performance.

Yang et al. [38] examined whether VR could support low achieving learners of English in improving their communication skills. Participants completed pre and post tests measuring listening comprehension and dialogic interaction. Post test mean scores were significantly higher than pre test scores, indicating that learners benefited from VR supported practice. However, the reported effect size was small (d = .30), suggesting that although VR had a positive impact, the magnitude of improvement was limited.

Xie et al. [37] compared the use of Cardboard based VR with a conventional PowerPoint presentation. Quantitative analyses showed that learners using VR produced presentations with significantly richer content and more appropriate vocabulary, reflecting strong effects (content: d = 1.24; vocabulary: d = .85). In contrast, pronunciation and fluency did not improve over the course of the intervention. This suggests that while VR may enhance lexical and content related aspects of communicative performance, it does not necessarily lead to improvements in real time speech production or prosodic accuracy.

Taken together, these studies indicate that VR can support the development of specific dimensions of communicative competence. However, its overall impact appears limited and may require complementary instructional approaches to achieve more robust and comprehensive learning gains.

4.5 Vocabulary

Twelve studies examined vocabulary learning using VR and reported inconclusive results. While most studies found positive effects, several reported weaker outcomes, with VR performing no better or worse than alternative methods. Generalizability is limited due to heterogeneous HMD technologies and comparison conditions. Overall, findings on vocabulary acquisition are mixed, and results were therefore clustered by comparison condition to identify meaningful patterns.

Hsu et al. [14] and Madini and Alshaikhi [25] examined vocabulary learning outcomes in VR scenarios without direct comparison to other instructional media. Both studies reported statistically significant post test gains, indicating that VR can support vocabulary acquisition. Reported effect sizes ranged from small (d = .49) to large (d = 1.10).

Four studies compared HMD based VR with computer based learning methods. The results were highly heterogeneous. While some studies reported higher vocabulary accuracy for learners in the VR condition, such as Legault et al. [22], others found no statistically significant differences, as reported by Lai and Chen [20], or even lower performance for VR learners, as observed by Cheng et al. [5] and Papin and Kaplan Rakowski [29]. Effect sizes ranged from small (d = .30) to large (d = 1.10), reflecting inconsistent evidence across studies.

Three studies compared HMD based VR with smartphone based learning applications. Findings were mixed. In some cases, learners in immersive VR conditions outperformed those using mobile or video based applications, with large effect sizes of up to d = 1.00 [33, 34]. In contrast, another study reported higher post test performance for the smartphone group, although these differences were not statistically significant [26].

The remaining three studies contrasted VR based instruction with traditional non digital learning methods such as printed materials or flashcards. Two of these studies reported significantly better outcomes for VR learners, with large effect sizes of up to d = .90 [2, 3]. By contrast, Hartfill et al. [13] found superior performance for an analog flashcard group in terms of vocabulary recognition and recall, yielding a very large effect size (d = 1.54).

Overall, the findings indicates that VR may support vocabulary learning in particular contexts, but its effectiveness varies considerably depending on instructional design, comparison media, and implementation characteristics.

5 Discussion

The findings of this systematic review provide a differentiated perspective on the potential of VR in second language acquisition. While the results confirm that VR can create highly engaging and authentic learning environments, its overall effectiveness varies considerably across language skill areas. The pedagogical value of this technology appears to depend less on the technology itself than on instructional design, the integration of scaffolding mechanisms, and the degree of learner interaction. Across the reviewed studies, the use of VR generally yielded positive but heterogeneous outcomes, with effect sizes ranging from small to very large. Although the majority of studies reported advantages for VR based instruction, these effects were not consistent across all learning domains or study contexts. This pattern further indicates that the effectiveness of VR based language learning varies substantially across different language skills and instructional contexts. The CEFR offers a useful analytical framework for interpreting these findings, as its multi level structure enables a more differentiated assessment of where VR based interventions appear most effective. Notably, the stronger effects observed in oral communication align with the CEFR's action oriented perspective, which conceptualizes learners as social agents engaged in goal directed communication. VR can support this perspective by providing authentic scenarios for spoken interaction across proficiency levels, ranging from basic exchanges at A1 or A2 to more complex discourse at C1 or C2.

5.1 Skill-Specific Implications

Below, the general findings of the review are succinctly interpreted in relation to practical language learning, structured by language skill domain and aligned with the CEFR.

Reading comprehension
VR environments proved particularly useful for improving reading comprehension when combined with instructional scaffolding or guided reading strategies. Studies that incorporated explicit support mechanisms reported higher comprehension scores and more effective information processing. These findings suggest that integrating VR with structured pedagogical approaches can foster strategic and cognitive dimensions of reading, especially at intermediate CEFR levels from B1 to B2, where learners are required to handle increasingly complex texts independently.

Listening comprehension
Studies on listening comprehension consistently demonstrated that multimodal input enhances learner performance. The availability of visual and contextual information supported both the comprehension and retention of auditory material. These findings align with cognitive theories such as dual coding and multimedia learning, which explain how the simultaneous processing of visual and verbal information can deepen understanding. From a CEFR perspective, VR appears particularly beneficial for learners at lower proficiency levels from A1 to B1, where listening comprehension is often constrained by limited linguistic resources and a strong reliance on contextual cues.

Writing performance
The reviewed studies revealed moderate effects of VR based learning on writing performance, depending on task characteristics and instructional design. In several cases, learners demonstrated improvements in content development, coherence, and linguistic variety. For learners at intermediate to advanced CEFR levels from B1 to C1, immersive writing tasks may support a deeper understanding of discourse structures by embedding writing activities in meaningful communicative contexts.

Speaking and communicative competence
VR showed considerable potential for fostering communicative competence, particularly with regard to fluency, content development, and vocabulary use. Learners benefited from authentic interaction opportunities provided by immersive environments, which closely reflect the CEFR descriptors for spoken interaction and production. However, pronunciation and prosodic control showed little improvement, indicating that while VR promotes confidence and communicative content, additional feedback mechanisms are required to address phonetic accuracy. Overall, these findings suggest that VR aligns well with the CEFR's emphasis on functional communication, but less strongly with its focus on phonological control.

Vocabulary acquisition
Findings related to vocabulary acquisition were the most inconsistent. While some studies reported substantial learning gains, others found that VR was no more effective than traditional instructional approaches. This variability likely results from differences in instructional design, learning duration, and the types of vocabulary

tasks employed. Context rich and interactive VR environments tended to produce stronger outcomes than isolated word learning activities. From a CEFR perspective, vocabulary learning through VR may be particularly advantageous at lower proficiency levels from A1 to B1, where concrete and situational vocabulary plays a central role and can be effectively supported through immersive contextualization.

5.2 Relation to Other Reviews and Overview Studies on the Research Topic

The findings of the present review are consistent with key results reported in prior synthesis studies on immersive VR in language education, while extending this body of research through a CEFR based analytical perspective. Across existing reviews, immersive VR is most consistently associated with positive effects on oral language skills, particularly speaking, often with medium to large effect sizes [32]. These effects are commonly attributed to affective mechanisms such as reduced speaking anxiety and increased learner confidence [9, 16, 32]. Listening comprehension likewise benefits from immersive and multisensory embedding, as contextual and visual cues support the processing and retention of auditory input [16, 27].

These patterns align closely with the present findings, which likewise identify speaking and listening as the skill domains most responsive to VR based interventions. Beyond oral skills, prior research emphasizes that learning gains are strongly contingent on pedagogical design and interaction. Collaborative VR scenarios, instructional scaffolding, and alignment with learning objectives have been shown to support writing, interaction, and retention, particularly at lower CEFR proficiency levels [11, 30, 31]. While some reviews report advantages across all language skills, these effects are consistently described as design dependent [27]. Finally, meta analytic evidence indicates that VR based language learning yields a statistically significant but small overall advantage over traditional instruction [32].

5.3 Methodological and Pedagogical Implications

From a methodological standpoint, the studies reviewed here highlight the need for greater standardization in experimental design and statistical reporting. Many investigations relied on small sample sizes, short intervention periods, and heterogeneous proficiency measures, which limits comparability and reproducibility. The absence of consistent alignment with a standardized framework such as the CEFR further complicates the interpretation of findings, as learner proficiency and task difficulty are often operationalized differently across studies.

Overall, the evidence suggests that VR should be conceptualized as a complementary rather than a replacement technology. Its greatest potential lies in supporting specific communicative and cognitive processes rather than functioning as a self contained learning environment. When aligned with CEFR descriptors, VR can be purposefully tailored to address specific skill requirements, such as vocabulary development at beginner levels or interactive speaking practice at upper intermediate levels. Educators should therefore prioritize learning tasks that are both proficiency appropriate and pedagogically meaningful.

Moreover, careful attention must be paid to cognitive load and learner autonomy. The highly immersive nature of VR can overwhelm learners if learning activities are not deliberately structured. Scaffolding, opportunities for reflection, and guided interaction can mitigate these challenges and promote deeper learning.

5.4 Limitations

Several limitations of this review should be acknowledged. The included studies varied widely in methodological rigor, intervention duration, and participant characteristics, and many lacked longitudinal data, which limits conclusions regarding long term language retention. In addition, the heterogeneity of VR technologies and comparison conditions restricts the generalizability of the findings.

Many studies relied on media comparison designs in which immersive VR was contrasted with alternative instructional formats that differed not only in medium but also in instructional design, interactivity, and learner engagement. Consequently, observed effects cannot be unequivocally attributed to the medium itself and may be influenced by confounding factors such as novelty effects, motivational differences, or variations in pedagogical quality. Furthermore, study quality was not systematically weighted in the interpretation of results, which may have affected the robustness of cross study comparisons.

Finally, the findings may be subject to publication bias due to the limited range of databases consulted. As the review includes studies published between 2010 and 2022, more recent technological developments and emerging longitudinal evidence are not fully represented.

6 Conclusion

By employing the CEFR as an integrated analytical framework, this review enables a pedagogically meaningful interpretation of VR effects across language skill domains and proficiency levels. Mapping empirical findings onto CEFR descriptors allows for the identification of stages of language development at which immersive VR appears particularly supportive, as well as contexts in which its effects remain limited or strongly design dependent. In this respect, the findings show substantial convergence with results reported in comparable systematic reviews and synthesis studies in the field.

Across these bodies of work, immersive VR is most consistently associated with positive effects in speaking and listening, where reductions in affective barriers and the availability of contextual and multimodal cues support oral communication. These patterns are mirrored in the present review and underscore the role of immersion, presence, and interaction as key mechanisms underlying learning gains.

Similarly, the present analysis aligns with existing reviews in showing that outcomes for reading, writing, and vocabulary learning are more heterogeneous. Effects in these domains appear to be highly contingent on instructional scaffolding, task structure, and learner proficiency level, particularly at intermediate stages of language development. Studies that embed VR within collaborative, strategy based, or guided instructional frameworks tend to report more stable learning gains, a pattern that is likewise evident in the findings synthesized here.

Overall, the review suggests that VR may offer added value for language education, although its effects are uneven across skills, learning contexts, and instructional designs. Consistent with prior reviews, VR does not consistently outperform alternative instructional media and in some cases yields comparable or less favorable outcomes. Its effectiveness therefore appears to depend less on the technology itself than on pedagogical integration, task design, and alignment with curricular objectives.

When examined through the lens of the CEFR, the use of VR for language learning aligns conceptually with the framework's action oriented perspective, as immersive environments can support situated interaction, negotiation of meaning, and socially embedded communication. This conceptual alignment echoes interpretations in other reviews, which similarly emphasize the suitability of VR for communicative and interaction oriented learning goals while cautioning against assuming uniform effects across all language skills. Accordingly, the educational potential of VR lies in complementing rather than replacing traditional instruction.

Several constraints limit the generalizability of the current evidence base, including heterogeneous technologies, short intervention durations, and a lack of longitudinal research. These limitations are likewise identified in other synthesis studies and point to persistent methodological challenges in the field. Future research should therefore adopt longitudinal and mixed method designs and more systematically align learning outcomes with CEFR proficiency levels. In sum, VR represents a promising but context sensitive tool whose educational value depends on careful pedagogical design, curricular integration, and evidence informed implementation.

References

1. Acar, A., Cavas, B.: The Effect of Virtual Reality Enhanced Learning Environment on the 7th-Grade Students' Reading and Writing Skills in English. *Malaysian Online Journal of Educational Sciences* 8(4), 22–33 (2020), https://www.mojes.net
2. Chen, C., Yuan, Y.: Effectiveness of virtual reality on Chinese as a second language vocabulary learning: Perceptions from international students. *Computer Assisted Language Learning* (online first) (2023), https://doi.org/10.1080/09588221.2023.2192770
3. Chen, C.-H., Hung, H.-T., Yeh, H.-C.: Virtual reality in problem based learning contexts: Effects on problem solving performance, vocabulary acquisition, and motivation of English language learners. *Journal of Computer Assisted Learning* 37(3), 851–860 (2021).
4. Chen, Y.-T., Li, M., Huang, C.-Q., Han, Z.-M., Hwang, G.-J., Yang, G.: Promoting deep writing with immersive technologies: An SVVR supported Chinese composition writing approach for primary schools. *British Journal of Educational Technology* 53(6), 2071–2091 (2022).
5. Cheng, A., Yang, L., Andersen, E.: Teaching language and culture with a virtual reality game. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, USA (2017).
6. Chien, S.-Y., Hwang, G.-J., Jong, M. S.-Y.: Effects of peer assessment within spherical video based virtual reality on EFL students' speaking performance and learning perceptions. *Computers & Education* 146, 103751 (2020).
7. Council of Europe: *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press (2001).
8. Council of Europe: *Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Companion Volume*. Council of Europe Publishing (2020).
9. Dhimolea, T. K., Kaplan Rakowski, R., Lin, L.: A systematic review of research on high immersion virtual reality for language learning. *TechTrends* 66(5), 810–824 (2022).

10. Dolgunsöz, E., Yildirim, G., Yildirim, S.: The effect of virtual reality on EFL writing performance. *Journal of Language and Linguistic Studies* 14(1), 278–292 (2018).
11. Esteves, J. R., Cardoso, J. C. S., Gonçalves, B. S.: Design recommendations for immersive virtual reality applications for English learning: A systematic review. *Computers* 12(11), 236 (2023).
12. Feng, B. J., Ng, L. L.: Facilitating writing performance of EFL learners via virtual reality: Immersion, presence, embodiment. *Frontiers in Psychology* 14 (2023).
13. Hartfill, J., Gabel, J., Neves Coelho, D., Vogel, D., Räthel, F., Tiede, S., Ariza, O., Steinicke, F.: Word Saber: An effective and fun VR vocabulary learning game. *Proceedings of the ACM* (online first) (2020).
14. Hsu, C.-C., Chen, Y.-L., Lin, C.-Y., Lien, W.-C.: Cognitive development, self efficacy, and wearable technology use in a virtual reality language learning environment. *Current Psychology* (online first) (2021).
15. Huang, H.-L., Hwang, G.-J., Chang, C.-Y.: Learning to be a writer: A spherical video based virtual reality approach to supporting descriptive article writing. *British Journal of Educational Technology* 51(4), 1386–1405 (2020).
16. Huang, X., Zou, D., Cheng, G., Xie, H.: A systematic review of AR and VR enhanced language learning. *Sustainability* 13(9), 4639 (2021).
17. Keller, T., Bolliger, K., Massie, B.: A French foreign language lesson for children with German as a second language in virtual reality: A case study. *Research Square* (online first) (2024).
18. Khodabandeh, F.: Exploring the applicability of virtual reality enhanced education on extrovert and introvert EFL learners' paragraph writing. *International Journal of Educational Technology in Higher Education* 19 (2022).
19. Kuhlemeier, H., van den Bergh, H.: Relationships between language skills and task effects. *Perceptual and Motor Skills* 86(2), 443–456 (1998).
20. Lai, K.-W. K., Chen, H.-J. H.: A comparative study on the effects of a VR and PC visual novel game on vocabulary learning. *Computer Assisted Language Learning* (online first) (2021).
21. Lee, A.: Using virtual reality to test academic listening proficiency. *Korean Journal of English Language and Linguistics* 19**,** 688–712 (2019).
22. Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., Li, P.: Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages* 4(1), 13 (2019).
23. Liberati, A., Altman, D. G., Tetzlaff, J., et al.: The PRISMA statement for reporting systematic reviews and meta analyses. *PLoS Medicine* **6**(7), e1000100 (2009).
24. Liu, C.-C., Guo, Y., Hwang, G.-J., Tu, Y.-F., Wang, Z.: Effects of an article structure strategy based spherical video virtual reality approach on EFL learners' reading comprehension. *Interactive Learning Environments* (online first) (2023)
25. Madini, A., Alshaikhi, D.: Virtual reality for teaching ESP vocabulary: A myth or a possibility. *International Journal of English Language Education* **5**, 111–123 (2017).
26. Nicolaidou, I., Pissas, P., Boglou, D.: Comparing immersive virtual reality to mobile applications in foreign language learning in higher education: A quasi experiment. *Interactive Learning Environments*, 1–15 (2021).
27. Özgün, O., Sadik, O.: Implementation of VR technologies in language learning settings: A systematic literature review. *Educational Policy Analysis and Strategic Research* 18(4), 32–61 (2023).
28. Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al.: The PRISMA 2020 statement. *BMJ* 372, n71 (2021).
29. Papin, K., Kaplan Rakowski, R.: A study of vocabulary learning using annotated 360 degree pictures. *Computer Assisted Language Learning* (online first) (2022).
30. Parmaxi, A.: Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments* 31(1), 172–184 (2023).
31. Peixoto, B., Pinto, R., Melo, M., Cabral, L., Bessa, M.: Immersive virtual reality for foreign language education: A PRISMA systematic review. *IEEE Access* 9, 48952–48962 (2021).
32. Qiu, X. B., Shan, C., Yao, J., Fu, Q. K.: The effects of virtual reality on EFL learning: A meta-analysis. Education and Information Technologies 29, 1379–1405 (2024).
33. Repetto, C., Di Natale, A. F., Villani, D., Triberti, S., Germagnoli, S., Riva, G.: The use of immersive 360 degree videos for foreign language learning. *Interactive Learning Environments*, 1–16 (2021).
34. Tai, T.-Y., Chen, H. H.-J.: The impact of immersive virtual reality on EFL learners' listening comprehension. *Journal of Educational Computing Research* 59(7), 1272–1293 (2021).
35. Tai, T.-Y., Chen, H. H.-J., Todd, G.: The impact of a virtual reality app on adolescent EFL learners' vocabulary learning. *Computer Assisted Language Learning* 35(4), 892–917 (2022).
36. Wang, Z., Guo, Y., Wang, Y., Tu, Y.-F., Liu, C.: Effects of a visual prompt scaffolding based virtual reality approach on EFL learners' reading comprehension. *Sustainability* 13(24), 13977 (2021).
37. Xie, Y., Chen, Y., Ryder, L. H.: Effects of mobile based virtual reality on Chinese L2 students' oral proficiency. *Computer Assisted Language Learning* 34(3), 225–245 (2021).
38. Yang, F.-C. O., Lo, F.-Y. R., Hsieh, J. C., Wu, W.-C. V.: Facilitating communicative ability of EFL learners via high immersion virtual reality. *Journal of Educational Technology & Society* 23(1), 30–49 (2020).
39. Yang, G., Chen, Y.-T., Zheng, X.-L., Hwang, G.-J.: From experiencing to expressing: A virtual reality approach to facilitating pupils' descriptive paper writing. *British Journal of Educational Technology* 52(2), 807–823 (2021).