Virtual Reality and Linguistic Output in Task-Based Foreign Language Speaking

Nami Takase
t.shizuoka11@gmail.com
Shizuoka University, Faculty of Informatics

Abstract

While earlier studies suggest that Virtual Reality (VR) can boost learner motivation and lower anxiety, its impact on language production has been less clear. This study examines how VR can be integrated with Task-Based Language Teaching (TBLT) to support second language (L2) learning. It compares the outcomes of a VR-based speaking task with those of a conventional videoconferencing tool (Zoom). The study particularly focuses on learners' vocabulary use by part of speech and their perceptions of task performance in a VR setting. Data were collected from 32 Japanese university students who performed a map-based speaking activity using VR and Zoom. The results of the study indicated that VR use was associated with higher word counts, more frequent use of pronouns, and positive learner perceptions of engagement and recall. At the same time, practical challenges remain regarding classroom implementation and long-term feasibility.

Keywords: *speaking activity, task-based language teaching, virtual reality*

Introduction

Task-Based Language Teaching (TBLT) emphasizes meaningful communication achieved through goal-oriented, authentic tasks, rather than on the explicit teaching of linguistic forms (Skehan, 1996, 1998, 2009). Recent technological developments have broadened the range of task environments available to teachers and learners, with Virtual Reality (VR) offering rich multimodal immersive contexts for interaction. A growing body of work suggests that VR can increase learner motivation, reduce speaking anxiety, and, under certain conditions, improve performance on communicative tasks (Chen, 2009). However, empirical studies also report the negative effects of VR on language practice. For example, the ecology of a VR environment may only partially approximate lived reality. Factors that disrupt language participation and learning include hurdles like interaction design and device usability, which impose non-trivial setup and operational demands, and cybersickness. In short, VR can function as both positive and negative in pedagogical practice (Kaplan-Rakowski et al., 2021).

This study investigated whether speech-focused TBLT in a VR environment yields measurable differences in second language (L2) learners' language production relative to a conventional two-dimensional (2D) platform.

Literature review

VR and task-based learning

VR environments provide learners with rich visual contexts, high levels of engagement, and opportunities for authentically situated communication (Huang et al., 2021; Shadiev & Yang, 2020). Within TBLT, these features are pedagogically important: by making objects and actions visible in a shared scene, VR can support message planning and allow learners to devote more attention to language use, focusing on how to say things rather than inventing content based on what comes to mind (cf. Skehan, 1996, 1998, 2009).

TBLT has been increasingly implemented in classrooms, emphasizing meaning-focused tasks with incidental attention to form. VR offers a multimodal information-rich medium that can reshape task conditions and provide new ways to scaffold communication. However, a persistent challenge in TBLT is the trade-off among Complexity, Accuracy, and Fluency (CAF). Learners have limited attentional resources, hence they cannot fully optimize all three aspects of language performance simultaneously in real time (Skehan, 1996, 1998, 2009). This study examines the efficacy of integrating task-based activities into actual classroom learning through VR, with the expectation that VR can alleviate CAF constraints by offloading conceptual demands via visual grounding, thereby enabling learners to focus more on linguistic performance and minimize trade-off effects. In addition, VR allows for the use of gestures in virtual spaces. Thus, it is hypothesized that VR can serve as an effective environment for second language practice.

Research has shown that an effective instructional design can enhance learner retention by optimizing the presentation of materials. According to the cognitive load theory, strategies that manage the complexity and interactivity of the material and increase the mental effort devoted to building and refining schemas support learning (Gkintoni et al., 2025; Paas et al., 2003). VR can help restore balance among learning components by presenting information within perceptually rich, spatially organized environments. When learners make full use of the affordances of a virtual space, less cognitive effort is spent constructing the task, allowing greater focus on language production. The head-mounted display also reduces external distractions, enabling more mental resources to be directed toward linguistic planning in task-based activities.

VR can also heighten motivation and engagement. It provides opportunities for worked examples, just-in-time scaffolding, and feedback that encourages active and reflective learning. However, these benefits depend strongly on instructional design (Gkintoni et al., 2025). The same immersive features that support learning can also increase extraneous cognitive load if tasks are overly complex, interfaces are unfamiliar, interactions are cumbersome, or sensory input is overwhelming, as in cases of cybersickness (Kaplan-Rakowski et al., 2021).

Therefore, thoughtful design is essential. Supportive measures such as pretraining in system controls, gradual

introduction of features, clear cues for task-relevant information, segmentation of complex sequences, and carefully calibrated realism can help reduce unnecessary load and direct learners' attention toward L2 learning goals.

Embodiment and gesture in VR

According to the embodied cognition theory, conceptual representations are grounded in bodily states and sensorimotor interactions, with cognition shaped through action and perception in the world (Barsalou, 1999; Reggin et al., 2023). Linguistic meaning is thus anchored not only in propositional content within VR, but also in learners' sensorimotor engagement with the shared scene, enabling abstract expressions to be connected to situated experiences. This dynamic is particularly significant in the context of L2 acquisition (Reggin et al., 2023).

Among co-speech gestures, deictic pointing is especially important because it helps speakers and listeners establish common ground. Pointing clarifies who or what is being referred to, directs attention, and aligns perspectives by tying expressions like "this," "there," or "turn left" to visible positions in space (Kita & Emmorey, 2023; Tellier & Ghio, 2021). In language learning, these gestures act as deliberate scaffolds: they ground reference, reduce ambiguity, and support both message planning and comprehension.

In VR environments, the value of deictic pointing becomes even more apparent. Because learners and interlocutors share the same virtual scene, gestures can mark precise locations, paths, or directions, helping prevent confusion when several possible referents appear in different areas. Similar deictic cues accompany pronouns in signed languages and in co-speech contexts, serving the same clarifying function (Kita & Emmorey, 2023). These resources are particularly useful for learners with limited vocabulary or those working through complex forms, as pointing supplements incomplete language with immediate spatial information. VR thus provides a rich set of embodied cues that learners can draw on when formulating speech.

Research also helps explain why gestures, especially deictic ones, play such a central role in L2 development. Spatial contexts support grounded meaning: when learners navigate a virtual street or explore a room, spatial expressions such as prepositions and locatives become tied to action and perception rather than remaining abstract symbols. This view is consistent with accounts that emphasize how conceptual categories develop through repeated sensorimotor experience (Barsalou, 1999).

Second, enactment enhances memory. Pairing verbal input with corresponding bodily action, for example, hearing "turn left" while pointing or reorienting one's head, strengthens encoding and supports durable recall. This enactment effect has been consistently observed in gesture-augmented learning (Macedonia & Knösche, 2011).

Third, affect and presence shape learning. When performed in immersive VR, gestures heighten learners' sense of presence and engagement. Because the affective qualities of context (valence and arousal) influence both immediate processing and long-term acquisition, emotionally engaging VR scenes combined with gestures may channel attention in ways that enhance retention (Snefjella & Kuperman, 2016).

Fourth, comprehension and production are inherently multimodal. Language use is inherently multimodal, with gestures, gazes, and actions co-expressing meaning alongside speech. Deictic pointing in VR highlights this multimodality, reducing the inferential burden during conceptualization and supporting more fluid and incremental language production (Özyürek, 2021).

Taken together, these considerations suggest that task design in VR can provide clear benefits for L2 speaking. Learners can draw on deictic expressions and gestures to establish shared references, encode language through enactment, and engage in emotionally meaningful interactions that are supported by multimodal cues. However, these benefits depend on instructional design. If navigation or interface control consumes too much attention, the sensorimotor richness that normally grounds meaning may generate an extraneous cognitive load. Therefore, the key issue is not simply whether embodiment supports learning but whether tasks and instructional design are appropriately structured. The effectiveness of VR can only be meaningfully measured when these conditions are considered in terms of their impact on language production.

3D environments

Empirical studies comparing three-dimensional (3D) or stereoscopic 3D (S3D) media with traditional 2D displays have yielded mixed results. Kaplan-Rakowski et al. (2021) found no compelling evidence that 3D formats inherently enhance vocabulary learning. In some cases, highly immersive 3D environments have actually reduced performance by increasing the extraneous cognitive load or diverting attention from the target material (Kaplan-Rakowski, 2019; Makransky et al., 2019). In a controlled study, learners were asked to study 16 Finnish nouns presented for 15 seconds each in either 2D or anaglyph S3D. The results revealed no learning advantage for the S3D group, and many participants reported discomfort consistent with cybersickness (Kaplan-Rakowski et al., 2021). The researchers suggested that the additional visual richness may have redirected attention away from lexical encoding toward the immersive display; the limited exposure time was insufficient to meet the processing demands of S3D; and the novelty and usability issues further contributed to distraction. They also noted that methodological constraints, such as the use of isolated concrete nouns, single exposures, and text-only assessments, limit the generalizability of these findings.

Building on this, comparisons of different levels of immersion reinforce the idea that "more immersive" is not always "more effective." Papin and Kaplan-Rakowski (2022) compared high-immersion VR (HiVR), low-immersion VR (LiVR) delivered on desktop monitors, and conventional 2D and found that LiVR with annotated 360° images outperformed both HiVR and 2D in vocabulary recognition, although the learners reported higher engagement in HiVR. Papin and Kaplan-Rakowski (2022) explained LiVR's advantage as its balance between providing contextualized spatial-semantic cues and minimizing the usability difficulties and discomfort often associated with head-mounted VR displays. Their results diverge from those of prior studies that have reported HiVR advantages (e.g.,Legault et al., 2019); this can be attributed to differences in media formats, such as 360° images versus 360° video, along with task design and the availability of multimodal supports, such as audio.

Present study and research questions

In summary, although prior research on immersive technologies has shown mixed results regarding language learning, little is known about how these dynamics unfold during real-time communication tasks. Studies comparing 2D and 3D media have suggested that VR immersion alone does not guarantee improved outcomes; instead, its effectiveness depends on how learners' attention, cognitive resources, and linguistic performance interact within the environment. Building on these studies, the present study employs a straightforward instructional design that minimizes task difficulty in 3D environments to elucidate the distinctive features of language activity in 3D contexts. Specifically, we examine how learners perform a map-based speaking activity in a fully immersive VR environment and a conventional Zoom setting. This study addresses the following two research questions:

RQ1. How does VR immersion influence the quantity of L2 output?

RQ2. What are the potential differences in part-of-speech distribution across the two modes?

Methods

Participants

This study involved 32 first-year Japanese university students (aged 19–21) enrolled in an English course. Their English proficiency levels ranged from A2 to B1 on the CEFR scale. Students were randomly divided into a VR (n = 18) and Zoom group (n = 14). Participation was voluntary and students were informed that the activity would not affect their course grades.

Procedure

A pre-/post-test design was employed to compare task-based speaking performance across the VR and Zoom groups, with both groups completing equivalent speaking tasks and a post-task survey. In the VR group, learners used the *Wander* application on Oculus Quest 2 or 3 headsets, and received technical support, orientation, and brief practice with a human assistant available throughout. Learners created private rooms, invited proficient English-speaking partners, and explored and discussed their favorite destinations in Japan

while navigating the virtual environment. The interlocutors appeared as avatars. Each learner repeated the task three times with different interlocutors. In the Zoom group, students were also paired with proficient English-speaking partners and spoke about their favorite places in Japan while using a shared map; pairs changed across three 15-minute sessions. Pre- and post-speaking tests were administered individually to both groups, wherein participants were asked to explain their favorite place in Japan for one minute. Materials for the VR group included Meta Quest 2/3 HMDs, controllers, stable Wi-Fi, the Meta Quest app, and the *Wander* application. Google Maps was used for the Zoom group. All speaking tests and task sessions were recorded, transcribed, and rated independently by two raters, and a post-task questionnaire was administered to capture the learners' perceptions of engagement and usability.

The survey was conducted after the post-test. The questionnaire included items on prior experience with VR, difficulties encountered during the task, symptoms of motion sickness, and perceptions of the advantages and disadvantages of VR.

Data analysis

All data were transcribed with fillers, disfluencies, repetitions, and Japanese segments removed. The transcripts were then annotated with part-of-speech tags using the Stanford POS Tagger in R. To ensure reliability, two raters independently checked 25% of the transcripts, yielding an inter-rater reliability of 90%. Discrepancies in POS tagging were jointly reviewed and corrected by the raters, whereas the remaining transcripts were individually verified by them. Total word counts were calculated using the statistical software R.

Results

The results revealed a notable increase in word count from pre-to post-test in both groups (Table 1). In the VR group (n = 18), the median word count rose from 54.50 (IQR = 13.00) to 72.00 (IQR = 15.20), with a Wilcoxon signed-rank test indicating a significant improvement ($Z = 1.63$, $p < .01$, $r = .39$; medium effect). Similarly, in the Zoom group (n = 14), the median increased from 49.00 (IQR = 26.50) to 75.00 (IQR = 19.80), showing a significant gain ($Z = 2.33$, $p < .01$, $r = .62$; large effect). These findings suggest that both VR- and Zoom-mediated tasks effectively promoted greater language output, with the Zoom group showing a slightly larger effect size.

Table 1. Word counts in the pre- and post-tests

		Pre-Test			Post-Test		Pre-test-Post-test comparison			
Mode	*n*	*Mdn*	*IQR*	*n*	*Mdn*	*IQR*	*p*	*Z*	*r*	
VR	18	54.50	13.00	18	72.00	15.20	0.00	1.63	0.39	**
Zoom	14	49.00	26.50	14	75.00	19.80	0.01	2.33	0.62	***

Note. *small $r = .10$, **medium $r = .30$, ***large $r = .50$ (Mizumoto & Takeuchi, 2011)

A Wilcoxon signed-rank test was conducted to compare the pre- and post-test performances for the use of nouns, verbs, pronouns, adjectives, and adverbs across the VR and Zoom groups.

For the VR group (n = 18), the results indicated a significant increase in pronoun use from pre-test (Mdn = 7.5, IQR = 3.5) to post-test (Mdn = 10, IQR = 1.8), representing a significant increase ($W = 24.5$, $Z = 2.25$, $p = .024$, $r = .53$; small-to-medium effect size). No significant differences were observed for adverbs ($p = .117$) or adjectives ($p = .063$).

In contrast, in the Zoom group (n = 14), pronoun use did not differ significantly ($W = 46.5$, $p = .706$, $r = –.10$) between the pre-test (Mdn = 5, IQR = 2) and post-test (Mdn = 5, IQR = 3.8). However, adverb use significantly increased from pre-test (Mdn = 5, IQR = 2.8) to post-test (Mdn = 7.5, IQR = 3.5) ($W = 91$, $p = .016$, $r = .65$; large effect). Similarly, adjective use increased from pre-test (Mdn = 2.5, IQR = 3.8) to post-test (Mdn = 4.5, IQR = 4.8) ($W = 89$, $p = .002$, $r = .82$; large effect). Pronoun usage frequency for both groups is shown in Figure 1.

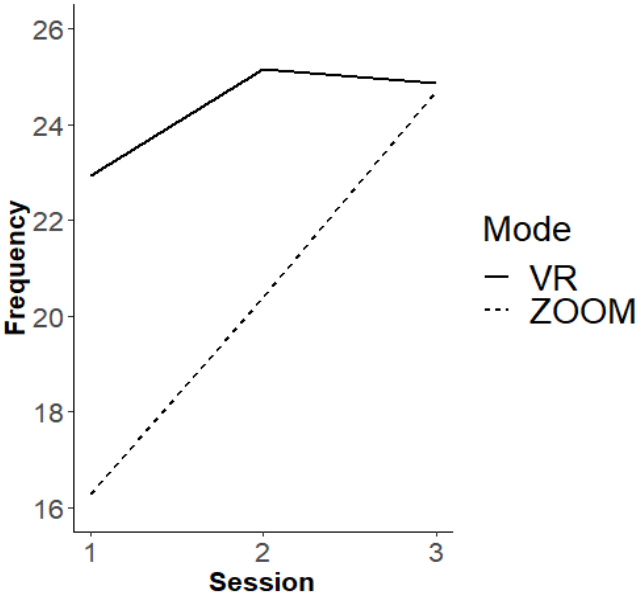

Figure 1. Pronoun usage frequency.

In summary, the VR group showed a significant improvement in pronoun use only (Figure 2), whereas the Zoom group demonstrated significant improvements in all parts of speech, including adverb and adjective use, but not in pronoun use (Figure 3).

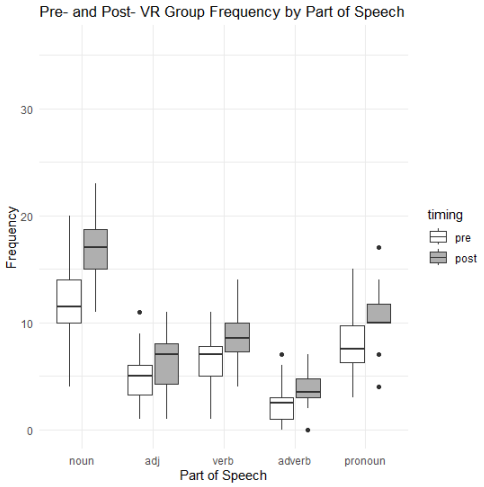

Figure 2. Pre- and post-VR group frequency by part of speech

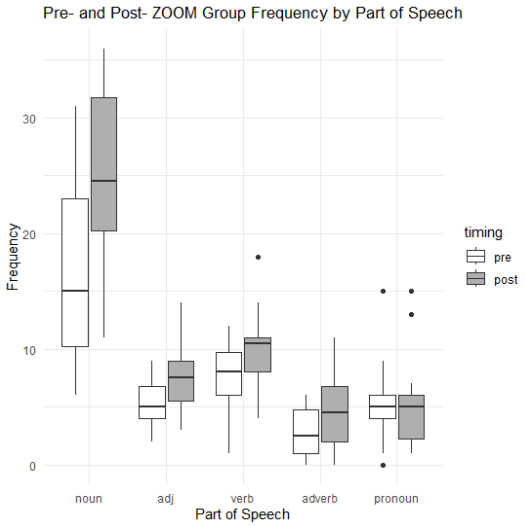

Figure 3. Pre- and post-Zoom group frequency by part of speech

Participants reported several types of difficulties encountered during the VR task. Eight respondents mentioned problems regarding VR operation and equipment use, such as difficulty moving smoothly, not knowing how to start moving first, and struggling with buttons, keyboards, or headset equipment. One participant noted that the recording function might not have worked properly. Five respondents highlighted challenges with English language use, including a lack of vocabulary, frequent pauses, difficulty translating items such as food names or signs, and overall struggles in speaking English. One respondent pointed out that choosing a location (a mountain) made navigation more difficult. Three other respondents stated that they had not encountered any problems. Overall, the primary difficulties were concentrated in two areas: VR operations (8/16 responses) and English expressions (5/16 responses).

Discussion

Regarding how VR immersion affects the amount of L2 output (RQ1), the task-based activity compared total word counts between the VR and Zoom groups. Both groups produced more language after the treatment than in the pre-test, suggesting that performing a task in VR does not necessarily impose a cognitive burden that interferes with learning. Although Kaplan-Rakowski (2019) noted that VR can introduce additional extraneous cognitive load, the present findings show that, while the effect size was somewhat larger for the Zoom group, the use of VR did not reduce overall language production. Post-task survey responses support this interpretation: eight of the sixteen participants reported that operating the VR system was difficult; however, this did not seem to prevent them from completing the task. In short, VR tasks enabled learners to increase their output compared with the pre-test.

For the differences in the distribution of parts of speech across modes (RQ2), participants in the VR group showed greater use of pronouns after completing the tasks, a pattern not observed in the Zoom group. Conversely, learners using Zoom increased their use of adjectives and adverbs from pre- to post-test, whereas the VR group showed little change in these categories. Across all three practice sessions, the VR participants consistently used more pronouns than those in the Zoom group. This result suggests that pronouns play a central role in VR-based communication, possibly helping speakers establish reference, reduce ambiguity, and organize their messages (Kita & Emmorey, 2023; Tellier & Ghio, 2021). Although the study did not track gestures, the findings imply that the immersive VR setting encourages learners to use pronouns as strategic tools for interaction and meaning-making.

Taken together, these findings indicate that VR and Zoom tasks facilitated similar overall levels of language production, but the distribution of linguistic resources differed in meaningful ways. In particular, the increased reliance on pronouns in VR suggests that learners adapted their communicative strategies to the affordances of immersive environments. The surrounding 360° space encouraged speakers to anchor their explanations in shared perspectives with their interlocutors, whereas Zoom participants tended to elaborate

their descriptions through adjectives and adverbs in the fixed screen-based view. This contrast underscores how task modality shapes not only the quantity of output but also the linguistic means through which learners construct meaning.

The results also reaffirm the importance of instructional design within task-based language teaching when integrating VR into classroom practice. As noted above, half of the participants reported difficulties with VR operation, even though they were paired with English-proficient partners who had been trained to provide technical support. While this arrangement allowed communication to proceed smoothly in the present study, scaling up to larger classes would present considerable challenges in securing an adequate number of trained facilitators. Moreover, effective implementation requires more than pedagogical planning; stable broadband capacity, regular device maintenance, and systematic management of head-mounted displays are prerequisites for sustainable use. Without such infrastructural support, the advantages of immersive VR tasks could be offset by logistical and technical difficulties.

In summary, although VR did not increase the total language output relative to Zoom, it shaped the distribution of parts of speech in ways that highlight the communicative potential of immersive contexts. This suggests that VR can enrich language learning by fostering interactional strategies such as pronoun use, provided that technical and institutional barriers are adequately addressed. At the same time, Zoom was also found to facilitate the use of adjectives and adverbs and allowed learners to convey more descriptive content. This underscores the need for careful consideration when selecting the mode of instruction to determine a suitable environment for specific pedagogical goals.

Conclusion

This study investigates how VR and Zoom environments influence L2 speaking performance in a map-based task. Both conditions resulted in significant increases in total word production, suggesting that immersive VR does not impose a cognitive burden strong enough to restrict language output. However, the linguistic resources learners employed differed across modalities. In VR, participants made greater use of pronouns, reflecting the shared spatial reference and embodied interaction that the immersive setting affords. By contrast, those in the Zoom group relied more on adjectives and adverbs, extending their descriptions through lexical elaboration rather than deictic reference.

These findings point to the importance of examining not only how much language learners produce, but also how they construct meaning within different communicative modes. From a TBLT perspective, VR appears to promote referential clarity and perspective-taking through increased pronoun use, supporting more interactive exchanges. However, Zoom tends to encourage greater descriptive precision through expanded use of modifiers. Therefore, choosing between these environments should align with instructional aims, whether the focus is on developing interactional grounding or enhancing descriptive fluency.

The findings also underscore the challenges of classroom implementation of VR. Technical difficulties, equipment management, and the need for facilitator support remain barriers to scaling VR beyond small group settings. Future research should investigate how these obstacles can be mitigated and explore how multimodal resources in VR, such as gestures, gaze, and movement, interact with language production over longer instruction periods.

Overall, VR and Zoom offer complementary advantages for language learning. By carefully aligning the task design and instructional objectives with the strengths of each mode, educators can leverage both immersive and conventional platforms to enrich learners' communicative competence.

Limitations and recommendations

This study has some limitations. First, the sample size was relatively small (n = 32) and drawn from a single university in Japan, which limits the generalizability of the findings to broader populations of L2 learners. Second, although technical support and orientation were provided, many participants reported difficulties in operating VR equipment. These operational challenges may have influenced task performance and perceptions, making it difficult to disentangle linguistic outcomes from technology-related factors. Third, the duration of the intervention was short, consisting of only three task sessions. Longer-term exposure might

yield different effects, particularly with respect to learner adaptation, reduction of operational difficulties, and sustained language development. Fourth, the study did not directly measure multimodal features such as gesture or gaze, which likely interacted with pronoun use and other communicative strategies in VR. Finally, infrastructural factors such as stable internet connectivity, headset maintenance, and classroom management of VR devices present significant challenges for scalability that were not fully addressed in this research.

Future research should extend this work in several directions. Expanding the sample to include learners from different proficiency levels, age groups, and educational contexts would allow for stronger claims of generalizability. Longer-term studies are also needed to investigate whether initial difficulties with VR operation diminish over time and whether continued use leads to sustained gains in communicative strategies. Incorporating multimodal analysis, such as capturing gesture, gaze, and body movement, would provide richer insights into how VR environments foster embodied interaction and support L2 development. From a pedagogical perspective, teacher training and structured pre-task familiarization with VR controls are recommended to minimize extraneous cognitive load and improve learner confidence. Institutions planning to implement VR should also invest in infrastructure, including reliable internet connections, systematic equipment management, and sufficient facilitator support. Finally, comparative studies can explore hybrid approaches that combine VR with videoconferencing or other technologies, aligning instructional mode with specific pedagogical objectives such as promoting referential clarity, scaffolding interaction, or enhancing descriptive fluency.

Acknowledgement

This study was supported by a Grant-in-Aid for Scientific Research (KAKENHI). The authors would like to express their sincere gratitude to the technical assistants who provided essential support with the VR equipment and software throughout the study. Special thanks are also extended to the student participants, whose active involvement in the VR activities made this research possible.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Chen, C. J. (2009). Theoretical bases for using virtual reality in education. *Themes in Science and Technology Education, 2*(1–2), 71–90.

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging cognitive load theory: The role of educational neuroscience and artificial intelligence in redefining learning efficacy. *Brain Sciences, 15*(2), 203. https://doi.org/10.3390/brainsci15020203

Huang, X., Zou, D., Cheng, G., & Xie, H. (2021). A systematic review of AR and VR enhanced language learning. *Sustainability, 13*(9), 4639. https://doi.org/10.3390/su13094639

Kaplan-Rakowski, R., Lin, L., & Wojdynski, T. (2021). *Learning vocabulary using 2D pictures is more effective than using immersive 3D stereoscopic pictures* (SSRN Scholarly Paper No. 3860239). SSRN. https://doi.org/10.2139/ssrn.3860239

Kita, S., & Emmorey, K. (2023). Gesture links language and cognition for spoken and signed languages. *Nature Reviews Psychology, 2*(7), 407–420. https://doi.org/10.1038/s44159-023-00186-9

Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., & Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages, 4*(1), 13. https://doi.org/10.3390/languages4010013

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education, 5*(4), 196–211. https://doi.org/10.1111/j.1751-228X.2011.01129.x

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning, 35*(6), 691–707. https://doi.org/10.1111/jcal.12375

Mizumoto, A., & Takeuchi, O. (2011). Introduction to effect size and statistical power: For accurate use of statistical analysis. *Bulletin of Chapter Reports, 2010*, 47–73.

Özyürek, A. (2021). Considering the nature of multimodal language from a crosslinguistic perspective.

Journal of Cognition, 4(1), 165. https://doi.org/10.5334/joc.165

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Papin, K., & Kaplan-Rakowski, R. (2022). *A study of vocabulary learning using annotated 360° pictures. SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3696821

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reggin, L. D., Gómez Franco, L. E., Horchak, O. V., Labrecque, D., Lana, N., Rio, L., & Vigliocco, G. (2023). Consensus paper: Situated and embodied language acquisition. *Journal of Cognition, 6*(1), 63. https://doi.org/10.5334/joc.308

Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability, 12*(2), 524. https://doi.org/10.3390/su12020524

Skehan, P. (1996). Second language acquisition research and task-based instruction. *Language Teaching, 29*(4), 189–211. https://doi.org/10.1017/S026144480000165X

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. Applied Linguistics, 30(4), 510–532. https://doi.org/10.1093/applin/amp047

Snefjella, B., & Kuperman, V. (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition, 156*, 135–146. https://doi.org/10.1016/j.cognition.2016.07.010

Tellier, M., Stam, G., & Ghio, A. (2021). Handling language: How future language teachers adapt their gestures to their interlocutor. *Gesture, 20*(1), 30–62. https://doi.org/10.1075/gest.19031.tel