

ARTICLE



Task-based language teaching in VR versus traditional settings

*Koyo Ogawa, Graduate School of Language & Literature,
Prefectural University of Kumamoto, Japan*

Abstract

This study compared Task-Based Language Teaching (TBLT) delivered in virtual reality (VR) and in traditional settings, examining learning gains, transfer, retention, and learner perceptions. 22 participants completed pretests, immediate posttests, transfer tests, and a questionnaire, while 10 of them, selected to balance proficiency across groups based on posttest scores, additionally completed a two-month delayed posttest and follow-up interviews. Wilcoxon signed-rank tests showed significant learning gains in both VR and traditional groups ($p < .001$; $p = .006$, in target discourse and $p = .004$; $p = .008$ in listening tests). ANCOVA indicated no immediate group difference in posttest scores ($p = .205$; $p = .322$) or in transfer measures ($p = .608$). After two months, a Mann–Whitney U test showed the VR group preserved target discourse significantly better than the traditional group ($p = .036$). Mann–Whitney U analyses of Likert-scale responses revealed greater enjoyment ($p = .008$), perceived retention ($p = .009$), reported difficulty ($p = .01$), and motivation ($p = .033$) in the VR group. Interview analyses (in vivo coding) highlighted VR affordances that promote contextualized practice and positive emotional engagement. These results suggest VR-TBLT may enhance long-term retention and learner engagement compared with traditional TBLT.

Keywords: virtual reality (VR); task-based language teaching (TBLT); transfer of learning; retention

Language(s) Learned in This Study: English

APA Citation: Ogawa, K. (2026). Task-based language teaching in VR versus traditional settings. *Language Learning & Technology*, 30(1), 1–20. <https://doi.org/10.64152/10125/73672>

Introduction

Task-Based Language Teaching (TBLT) is widely recognized as an effective approach for promoting meaningful second language use through authentic, goal-oriented tasks (Ellis, 2003; Khezrlou, 2025; Taguchi & Zhao, 2025). By requiring learners to use the target language to complete tasks, TBLT fosters communicative abilities that are transferable to real-world contexts (Nolen, 2020).

Virtual reality (VR) has recently attracted increasing attention in language education due to its affordances of immersion, presence, and interactivity (e.g., Taguchi & Zhao, 2025; Yan & Lowell, 2024). By placing learners in realistic and engaging environments, VR can make language use purposeful, potentially reducing affective barriers and enhancing motivation. Previous studies have reported positive effects of VR on language learning outcomes such as vocabulary acquisition and speaking fluency (e.g., Akyıldız, 2025; Albayrak et al., 2023).

Despite this growing interest, research that rigorously examines the integration of TBLT and VR remains limited (Taguchi & Zhao, 2025). In particular, few studies have directly compared VR-based TBLT (VR-TBLT) and traditional TBLT under controlled conditions using identical tasks, and even fewer have investigated transfer of learning or delayed memory retention. These dimensions are crucial for evaluating the pedagogical value of VR beyond short-term gains.

To address these gaps, the present study compares VR-TBLT and traditional TBLT using identical ice-cream ordering tasks with 22 participants. Learning outcomes were assessed through a pretest, immediate posttest, two-month delayed posttest, and a real-world transfer task, supplemented by questionnaire and interview data. By extending analysis beyond immediate performance to include transfer and delayed retention, and by qualitatively examining learners' experiences, this study provides insight into not only whether VR enhances task-based learning, but also how and why immersive contexts may support durable and transferable language use.

Literature Review: Integrating TBLT and Virtual Reality

Among the several definitions of “tasks” in TBLT, three defining characteristics are central: a primary focus on meaning, authenticity of learning materials, and the production of non-linguistic outcomes (Taguchi & Zhao, 2025). TBLT engages learners in goal-directed communication without prescribing specific grammatical forms, thereby promoting meaningful interaction. Taguchi and Zhao explain authenticity as “a close correspondence between tasks and real-life activities” (p. 3). Tasks also foster not only linguistic competence but also peripheral skills such as decision-making and problem-solving necessary for real-life contexts.

VR can strengthen these core characteristics of TBLT through its distinctive affordances—immersion, presence, and interactivity (e.g., Taguchi & Zhao, 2025; Yan & Lowell, 2024; Yuditseva, 2023). Immersion, achieved through head-mounted displays (HMDs), isolates users from the physical world and heightens the sense of presence, or the feeling of “being there,” which enhances engagement and attention. Interactivity enables embodied learning: users can manipulate virtual objects, navigate spaces, and perform actions that align with linguistic functions. These embodied interactions can link language with physical experience, thereby supporting deeper cognitive encoding and recall (Smith & McCurrach, 2021; Taguchi & Zhao, 2025).

Given these affordances, researchers increasingly view VR and TBLT as mutually reinforcing. VR's context-rich environments can make TBLT tasks more authentic, situating language learning in realistic communicative contexts (Taguchi & Zhao, 2025; Yan & Lowell, 2024). Such synergy may strengthen connections between language and context, enhance motivation, and promote retention. However, despite this theoretical alignment, relatively few empirical studies have fully implemented TBLT within VR settings. In their review of 33 studies, Taguchi and Zhao (2025) found that only 12 of 39 VR-based “tasks” met all three defining TBLT criteria, suggesting that much existing work remains task-like rather than genuinely task-based. Without grounding in TBLT principles, VR runs the risk of being reduced to a novelty tool that prioritizes technological excitement over pedagogical rigor.

Empirical studies have nonetheless reported encouraging outcomes. Chen and Liao (2021), though not meeting all TBLT criteria, found that participants using 360-degree VR workplace scenes outperformed a PowerPoint-based control group in listening, vocabulary, and conversational skills, and also reported higher motivation and satisfaction. Similarly, Chen et al. (2019) used Google Earth VR to support writing preparation activities, improving learners' expository writing even without a control group. Khodabandeh (2022) also observed superior writing performance in VR conditions while maintaining adherence to TBLT principles. These studies collectively suggest that VR-enhanced tasks can promote productive skills when pedagogical design is consistent with TBLT.

VR has also been used in task-like oral communication activities. Xie et al. (2019) examined twelve learners' performance across multiple tour-guiding simulations and found improvement in content and vocabulary, though not in fluency or grammar. Their results highlight the need for further investigation into which domains of oral communication benefit most from VR tasks. Overall, even studies meeting TBLT's formal criteria rarely articulate their theoretical rationale, limiting understanding of why such designs promote learning.

In response, there is growing consensus that VR-based language research must be pedagogically grounded, explicitly integrating TBLT's core principles rather than relying on superficial task-like activities. The combination of TBLT's pedagogical rigor and VR's immersive affordances holds promise for advancing language education.

Building on this gap, the present study systematically applies TBLT principles within immersive VR (VR-TBLT), strictly adhering to the three core task criteria—meaning orientation, task authenticity, and non-linguistic outcomes (Taguchi & Zhao, 2025). An ice-cream ordering task was selected for its clear communicative purpose, context-dependent decision-making, and everyday relevance. Pedagogical tasks such as information-gap and role-play remained meaning-oriented, requiring learners to use target expressions to achieve concrete goals while engaging non-linguistic skills such as decision-making, simple calculation, cultural interpretation, and collaboration. Details of the task design are presented later.

The study then compares VR-TBLT and traditional classroom-based TBLT across four dimensions: learning outcomes, real-world transfer, long-term retention of target discourse, and learners' perceptions. This design clarifies not only whether VR-TBLT is effective, but how its affordances shape learning, memory, and engagement relative to traditional TBLT.

Current Study

This study employs quasi-experimental mixed-methods design with both quantitative and qualitative approaches to compare participants' performance and perceptions in VR-TBLT and traditional TBLT, with the aim of clarifying the effectiveness of VR-TBLT. The quantitative analyses focus on learning outcomes, transfer of learning, and retention of target discourse, while the qualitative analyses explore participants' perceptions of their learning experiences in both conditions and provide deeper insights into what makes VR-TBLT distinct from non-VR settings. Accordingly, this study is guided by the following research questions:

1. How effective is VR-TBLT in promoting language learning compared to traditional TBLT, including its impact on real-world transfer and long-term retention?
2. How do learners' perceptions of their experiences differ between VR-TBLT and traditional TBLT?

Methodology

Participants

Twenty-two participants (16 female, 6 male; aged 18–27, $M = 21.5$) took part in the study. Their estimated CEFR levels ranged from A1 to B1 (15 A1, 2 A2, 5 B1), based on declared official scores (15 participants) or the [EF Standard English Test](#) (7 participants). Participants were recruited through university announcements and social media, and included both university students and non-student adults. They were assigned to either a VR (experimental) group or a non-VR (control) group based on preference and self-reported tolerance for VR-induced dizziness, resulting in a quasi-experimental design. The VR group ($n = 12$; 7 female, 5 male; $M = 21.3$) included 6 A1, 2 A2, and 4 B1 learners, while the control group ($n = 10$; 9 female, 1 male; $M = 21.9$) consisted of 9 A1 and 1 B1 learner.

Materials

Virtual Environment

Standalone PICO 4 VR devices with controllers and tracking technology were used, allowing full immersion without external computers. A fictional ice cream shop was developed using Blender and Unity and uploaded to VRChat, enabling multi-user interaction with custom avatars and worlds. Free 3D assets from the Unity Asset Store were used for general furniture, while the author created key items such as the ice cream display, menu, currency, and flavors. The resulting environment appears in [Figure 1](#).

Figure 1

A VR Ice-Cream Shop Created in VRChat



Real-World Café

A physical café was constructed at the university for transfer tests (Figure 2), including a counter, printed and electronic menus, coffee equipment, props, and US dollar bills. To maintain consistency across participants and ensure controlled testing conditions, two native English-speaking servers followed a standardized set of target discourse items (e.g., *For here or to go?*, *What can I get for you?*). While the servers' utterances were standardized, participants freely selected their drinks, customizations, and payment amounts, resulting in varied interactional paths. Details regarding the evaluation criteria and rationale for these design choices are provided in the Analysis section.

Task Handouts

In the learning process, several task handouts were distributed to participants. The control group got physical handouts while the experimental group saw virtual handouts in VR.

Tests, Questionnaires, and Interviews

Participants provided demographic and proficiency data via a brief questionnaire; those without official test scores completed the EF SET online test. Assessments included: (a) Pre- and post-treatment tests: target discourse knowledge and listening comprehension, (b) immediate post-treatment real-world transfer test and perception questionnaire, and (c) two-month delayed oral retention test and follow-up interviews.

To ensure content validity, all test items were developed directly from the target expressions taught during instruction. Drafts of the target discourse cloze test and listening comprehension test were reviewed by two native English-speaking experts, who evaluated their naturalness, sociopragmatic appropriateness, and contextual fit. Revisions were made to enhance authenticity and clarity. Likewise, the perception questionnaire and interview prompts were refined through graduate seminar discussions to ensure conceptual relevance and alignment with the research questions. Although statistical reliability indices were not available, expert review and iterative revision provided primary validity evidence.

Figure 2*A Real Café for Transfer Tests*

Procedure

The study comprised four phases: pre-task, treatment, immediate post-task, and delayed post-task. Procedures were identical across groups except for the delivery mode (VR vs. classroom). Tasks were completed within one or two days depending on availability, with no observed differences in fatigue or engagement.

Pre-Task Phase

Participants first took two pretests: a target discourse test and a listening comprehension test. The target discourse test measured prior knowledge of expressions used in server-customer interactions at an ice cream shop through eight fill-in-the-blank items based on Japanese prompts (Appendix A). The listening test presented a recorded ice cream order followed by six multiple-choice questions about the server's recommendations, the customer's choices, and the total bill.

Treatment Phase

Participants completed seven pedagogical tasks (PT1–PT7), described later. The VR group performed them in a virtual ice cream shop using head-mounted displays (HMDs), while the control group completed parallel activities in a classroom. To reduce fatigue, tasks were divided into three blocks (PT1–3, PT4–6, PT7) with breaks between them. Participants received immediate feedback on their responses upon completing each task, and their answers during this phase were not recorded for analysis. Although most participants finished in one day, total duration varied slightly due to logistical factors—for instance, the control group wrote answers on paper, while the VR group manipulated virtual objects and props. Listening comprehension tasks used identical pre-recorded audio with fixed pauses, ensuring consistent timing. The entire treatment lasted about five hours for the control group and just over six hours for the VR group. Although both groups completed the same pedagogical tasks, this additional exposure time for the VR group should be considered as a potential confounding factor when interpreting group differences, particularly in the retention outcomes.

Immediate Post-Task Phase

After the treatment, participants completed four measures: (1) a target discourse posttest, (2) a listening comprehension posttest, (3) a real-world transfer test, and (4) a reflective questionnaire. The first two mirrored the pretests. The transfer test evaluated participants' ability to apply learned language in an authentic café where they ordered drinks using U.S. currency within a \$6 budget. Although the native English-speaking servers followed a set of target expressions for consistency, they otherwise interacted naturally with participants according to their individual orders and customization choices. All performances were video-recorded for subsequent analysis. The questionnaire (eight 6-point Likert items and two open-ended) assessed enjoyment, motivation, concentration, memory, usefulness, and willingness to continue. Details of the scoring procedures and analysis methods are provided in the Analysis section.

Delayed Post-Task Phase

Two months later, five participants from each group completed a delayed target discourse test and a follow-up interview. Because the two groups' score levels were not comparable at the immediate post-task phase, the top five participants from the control group were first selected. Importantly, the sixth-highest control participant's score was substantially lower than that of any participant in the experimental group, making five the maximum number of participants whose proficiency could be fairly matched across groups. Subsequently, five experimental group participants with the most similar scores to the selected control group participants were chosen to ensure comparability (Welch's $t = .857$). While this selection procedure may introduce a potential bias, it allowed for a fair comparison between the groups. The delayed test, administered online via video calls, used the same items but required oral responses to assess memory retention. Interviews explored enjoyable aspects of learning experience and factors contributing to strong long-term recall.

Pedagogical Tasks (PTs)

The TBLT framework comprised three task types: input-based (PT1–3), information gap (PT4–6), and role-play simulation (PT7). Participants used target language to complete each task in authentic, real-life scenarios involving decision making, calculation, and communication. These PTs thus met all three task criteria: meaning orientation, authenticity, and non-linguistic outcomes.

Table 1 summarizes the seven pedagogical tasks (PT1–PT7), including task type, focus, expected outcomes, delivery mode, and approximate duration. Detailed descriptions are provided in the subsections that follow.

Input-Based Tasks

PT1: Target Discourse Activity

Participants listened to ice-cream ordering utterances and selected appropriate responses. The task had two parts: choosing customer responses to server prompts (e.g., *For here or to go?* → *To go, please*) and choosing server responses to customer utterances. Each part included six target expressions. The control group completed the task on paper, while the VR group selected answers by moving labeled virtual cubes in the virtual shop (Figure 3). The same audio was replayed for answer checking. The task lasted approximately 20 minutes.

PT2: Ice Cream Customization Activity

Participants listened to seven customer–server conversations and identified the correct ice cream orders, which increased in complexity (from a single scoop to double scoops with toppings and syrup). Twelve illustrated options were provided, including distractors.

The control group marked answers on a printed menu, whereas the VR group assembled virtual ice cream. Identical audio was played twice. This activity took about 20 minutes.

PT3: Payment Recognition Activity

Participants listened to seven orders and calculated correct payment amounts using a menu. The control group selected printed currency images, while the VR group used virtual bills and coins. Longer pauses were included to allow time for virtual manipulation. The activity lasted about 20 minutes.

Table 1

Overview of the Pedagogical Tasks (PT1–PT7)

Task	Type	Main Focus	Key Linguistic / Non-linguistic Outcome	Mode Difference (VR vs. Non-VR)	Approx. Duration
PT1	Input-based	Recognizing target discourse	Identifying appropriate customer/server responses	VR: select virtual cubes; Non-VR: paper	~20 min
PT2		Understanding customization language	Identifying correct ice-cream orders	VR: manipulate virtual ice cream; Non-VR: menu sheet	~20 min
PT3		Payment & calculation	Determining correct amounts using menu	VR: virtual money; Non-VR: printed currency	~20 min
PT4	Information gap	Sociocultural practice (tipping)	Exchanging tipping info; decision-making	VR: exchange virtual items; Non-VR: paper	20–30 min
PT5		Recommendations	Asking/giving suggestions; adjectives	Setting only (VR café vs classroom)	~20 min
PT6		Order customization	Asking/answering to complete partner's order	Setting only (VR vs classroom)	~40 min
PT7	Role-play	Full ordering performance	Complete interaction; decision-making, budget, tipping	VR: full object manipulation; Non-VR: imagination	20–30 min

Information Gap Tasks

PT4: Tipping in an Ice Cream Shop

Participants decided tipping amounts for six orders (\$3–\$8) and completed a paired information-gap role-play to exchange and record this information (Appendix B). Both groups used the same handout; only the VR group exchanged virtual ice cream and money. The task took about 20 minutes for the control group and 30 minutes for the VR group due to additional physical actions.

PT5: Ice Cream Recommendation

Working in pairs, participants exchanged recommendations using different menus (Handouts A and B) (Appendix C), asking for suggestions based on descriptors (e.g., *sweet, fruity*). The VR group interacted in the virtual shop, and the control group in the classroom. Since no object manipulation was required, both groups completed the activity in about 20 minutes.

PT6: Customizing Orders

Using different handouts (Appendix D), pairs exchanged information to identify complete orders (e.g., for here/to go, container, flavors, toppings), completing five exchanges per person. The task provided semi-controlled practice of the full ordering process. Each group took approximately 40 minutes to complete the task.

Figure 3

A Participant Placing Items to Express Her Answers



Role-Play Simulations

PT7: Complete Ordering Task

In the final open-ended simulation, participants enacted full ice cream orders, alternating between customer and server roles. They followed guidelines requiring them to order recommended flavors, include double scoops with toppings or syrup, tip at least 50 cents, and stay within budget (Appendix E).

The control group conducted the role-play in a classroom without props, relying on imagination. The VR group performed in the virtual ice cream shop, interacting with virtual objects such as ice cream, menus, and money. This realistic manipulation extended task duration to about 30 minutes, compared to 20 minutes for the control group.

The longer VR duration was primarily due to physical actions (e.g., handling virtual objects), not differences in pedagogical content or linguistic demands, and is therefore unlikely to have affected learning outcomes.

Analysis

Multiple methods were employed to address the two research questions. To address the first research question—To what extent is VR-TBLT effective in language learning compared to traditional TBLT?—participants' gains from pretest to immediate posttest were analyzed for both the target discourse and listening comprehension assessments ($N = 22$). Because the data were not normally distributed, within-group gains were examined using the Wilcoxon Signed-Rank test. To compare improvement between

groups while controlling for baseline differences, ANCOVA was conducted with pretest scores as covariates.

For the target discourse test, Levene's test indicated a slight violation of homogeneity of variances ($F = 4.524, p = .046$), but regression slopes were homogeneous and residuals were approximately normally distributed based on Q–Q plot inspection. Given ANCOVA's robustness to modest assumption violations with comparable sample sizes ($n = 12$ vs. $n = 10$), the analysis was deemed appropriate. For listening comprehension, no violation of homogeneity was found ($F = 2.607, p = .122$), and residuals were normally distributed, further supporting the use of ANCOVA.

Transfer of learning and long-term retention were also analyzed. Performance on the real-world transfer test ($N = 22$) was evaluated using a task performance rubric and compared between groups using the Mann–Whitney U test and ANCOVA. The rubric comprised seven criteria assessing target discourse use, communication skills, and task completeness, each rated on a 0–4 scale (Appendix F). Following East (2021), this analytic approach captures task performance more effectively than dichotomous measures.

All performances were rated by the author, and three randomly selected samples (3 participants \times 7 criteria) were independently scored by two additional experienced native-English educators. Inter-rater reliability, assessed using ICC (3,1), was good with Rater A (ICC = .757, 95% CI [.491, .894]) and good to excellent with Rater B (ICC = .837, 95% CI [.642, .931]), indicating acceptable reliability. For the transfer-test ANCOVA, Levene's test showed no violation ($F = 1.043, p = .319$), and residuals were approximately normal. Delayed posttest scores ($n = 10$) were analyzed using the Mann–Whitney U test after confirming group equivalence at the immediate posttest via Welch's t-test ($p = .857$).

The second research question—How do learners' perceptions of their learning experience differ between VR-TBLT and traditional TBLT?—was addressed using data from the exit questionnaire ($N = 22$) and follow-up interviews ($N = 10$). Participants' perceptions were analyzed quantitatively via Likert-scale responses using the Mann–Whitney U test and qualitatively via open-ended questionnaire responses processed in KH Coder. KH Coder identified frequently used words, guiding the design of follow-up interviews. Two semi-structured questions were used:

1. What did you find fun or enjoyable about learning in VR-TBLT/TBLT?
2. What do you think helped you retrieve the target discourse you learned two months ago during the delayed posttest? (asked only to participants scoring above 20/32)

Interview responses were analyzed using in vivo coding to preserve participants' original wording while minimizing researcher bias (Saldaña, 2016). Because coding was conducted by a single researcher (the author), several steps were taken to enhance trustworthiness. First, all English-translated comments and their corresponding in vivo codes are provided in the appendices to ensure transparency and traceability (Appendices G–H). Coding proceeded iteratively: initial line-by-line in vivo codes were generated and then grouped into preliminary categories based on shared meanings.

For the first interview question (enjoyable aspects of VR-TBLT/TBLT), categorization was informed by established features of TBLT (e.g., meaning orientation, authenticity, non-linguistic outcomes) and VR learning (e.g., interactivity, immersion, sense of presence), while additional categories were inductively created as needed. For the second question (factors supporting long-term recall), all categories and themes were derived inductively without relying on a predefined framework. Categories and themes were refined through repeated comparison to enhance analytical credibility.

Results

Immediate Learning Outcomes

Table 2 presents pretest and posttest scores for both groups ($N = 22$) on the target discourse and listening comprehension measures.

Table 2*Average Score Improvements from Pretests to Posttests*

	Experimental Group (n=12)	Control Group (n=10)
Target Discourse Pretest	15.38 / 32	8.18 / 32
Target Discourse Posttest	29.29 / 32	22.50 / 32
Listening Pretest	5.5 / 7	4.2 / 7
Listening Posttest	6.5 / 7	6.0 / 7

Wilcoxon Signed-Rank tests revealed significant gains in target discourse for both the experimental group ($p < .001$, $r = 1.00$) and the control group ($p = .006$, $r = 1.00$), indicating strong within-group effects. The rank-biserial correlations reached ± 1.00 because all non-zero differences reflected improvement, with no score decreases.

Because pretest scores differed substantially between groups, ANCOVA was conducted to examine relative improvement while controlling for baseline differences (Table 3).

ANCOVA showed that pretest scores significantly predicted posttest performance, $F(1, 19) = 9.90$, $p = .005$, partial $\eta^2 = .343$, whereas group membership was not a significant predictor, $F(1, 19) = 1.73$, $p = .205$, partial $\eta^2 = .083$. Thus, although both groups improved, VR-TBLT did not yield significantly greater gains than traditional TBLT after controlling for initial proficiency.

A similar pattern was observed for listening comprehension. Wilcoxon tests indicated significant improvements for both the experimental group ($p = .004$, $r = 1.00$) and the control group ($p = .008$, $r = 1.00$), again reflecting uniformly positive change. However, ANCOVA revealed no significant between-group difference in gains, $F(1, 19) = 1.04$, $p = .322$, partial $\eta^2 = .052$, while pretest scores strongly predicted posttest outcomes, $F(1, 19) = 55.49$, $p < .001$, partial $\eta^2 = .745$ (Table 4).

Table 3*ANCOVA Summary for Target Discourse Improvements*

Source	Sum Sq	df	F	p-value	partial η^2
Group (VR vs. Non-VR)	35.44	1	1.73	0.205	0.083
Pretest	203.22	1	9.90	0.005	0.343
Residual	390.01	19			

Table 4*ANCOVA Summary for Listening Comprehension Improvements*

Source	Sum Sq	df	F	p-value	partial η^2
Group (VR vs. Non-VR)	0.29	1	1.04	0.322	0.052
Pretest	15.64	1	55.49	< 0.001	0.745
Residual	5.36	19			

Transfer of Learning

Mean scores on the real-world transfer test were 24.92 (out of 28) for the experimental group and 21.30 for the control group. Although the experimental group scored higher, the Mann–Whitney U test showed that this difference was not statistically significant ($U = 84.00, p = .119, r = .400$), indicating no clear advantage of VR-TBLT over traditional TBLT in promoting transfer of learning.

Because the transfer test closely reflected the target discourse task, target discourse posttest scores were entered as a covariate in a follow-up ANCOVA. Results showed that posttest scores significantly predicted transfer performance, $F(1, 19) = 8.84, p = .008$, partial $\eta^2 = .317$, whereas group membership did not, $F(1, 19) = 0.27, p = .608$, partial $\eta^2 = .014$.

Although the between-group difference did not reach statistical significance, the observed mean gap (approximately 3.6 points on a 28-point scale) suggests a potentially meaningful practical advantage for the VR-TBLT group. Given the small sample size and limited statistical power, this result should be interpreted cautiously and warrants further investigation with larger samples.

Target Discourse Memory Retention

Table 5 presents delayed posttest scores for 10 selected participants, compared with their immediate posttest scores. Participants in the experimental group achieved a higher mean score on the delayed posttest ($M = 27.65$) than those in the control group ($M = 13.35$). A Mann–Whitney U test confirmed that this difference was statistically significant ($U = 23.0, p = .036, r = .840$), indicating superior long-term retention of target discourse in the VR-TBLT group.

This analysis was conducted on a small, matched subsample ($N = 10$), selected to ensure comparable immediate posttest performance across groups. Although this sampling procedure limits generalizability, it allows for a controlled exploratory comparison of long-term retention. The findings should therefore be interpreted with caution, but they nonetheless provide suggestive evidence that VR-TBLT may support stronger memory retention than traditional TBLT.

Table 5

Each Participant's Score in the Delayed Target Discourse Test

Experimental Group (VR)	Posttest Scores	Delayed Test Scores	Control Group (NON-VR)	Posttest Scores	Delayed Test Scores
Participant A	32	32	Participant F	32	28
Participant B	31.25	31.25	Participant G	32	7.75
Participant C	29.75	30	Participant H	27.75	9
Participant D	26.25	24.25	Participant I	27.25	11
Participant E	24.5	20.75	Participant J	26.5	11
Mean	28.75	27.65	Mean	29.1	13.35

Perceptions

Quantitative analyses of Likert-scale responses and KH Coder outputs reveal overall group-level trends, while in vivo coding provides insight into learners' individual experiences.

Overall Trends

Based on responses to eight 6-point Likert-scale items, Mann–Whitney U tests were conducted to examine between-group differences (Table 6).

Table 6*Mann-Whitney U Test Results for Questionnaire Responses*

Question	VR-TBLT (Mean)	Traditional TBLT (Mean)	U Value	p-value	r
Q1 (Satisfaction)	5.8	5.3	72.0	0.355	- 0.20
Q2 (Enjoyment)	6.0	5.4	90.0	0.008	- 0.58
Q3 (Motivation Changes)	5.6	5.3	70.0	0.481	- 0.16
Q4 (Concentration)	5.4	5.4	62.0	0.908	- 0.03
Q5 (Memory)	5.6	4.6	96.5	0.009	- 0.60
Q6 (Usefulness)	5.8	5.3	70.5	0.418	- 0.17
Q7 (Difficulty)	2.2	3.7	21.5	0.010	+ 0.64
Q8 (Desire to Continue)	5.8	5.1	88.0	0.033	- 0.46

Four items showed statistically significant differences. The VR-TBLT group reported significantly higher enjoyment (Q2, $p = .008$, $r = .58$), perceived memory support (Q5, $p = .009$, $r = .60$), and desire to continue learning (Q8, $p = .033$, $r = .46$). They also rated the learning experience as significantly more difficult than the traditional TBLT group (Q7, $p = .010$, $r = .64$). No significant differences were found for satisfaction, motivation change, concentration, or usefulness.

Because the questionnaire analysis was exploratory and the sample size was small, no correction for multiple comparisons was applied; instead, effect sizes were reported to support interpretation.

Open-ended questionnaire responses were analyzed using KH Coder. The VR-TBLT group produced substantially more text overall (1,592 words; 388 unique) than the control group (631 words; 186 unique). On average, VR-TBLT participants wrote more sentences and more than twice as many words per person, suggesting greater engagement in reflecting on their learning experiences.

Word frequency analysis showed that both groups frequently described the experience as fun or enjoyable. In the VR-TBLT group, *fun/enjoyable* ranked fourth (13 occurrences), whereas it ranked second in the control group (8 occurrences), indicating that enjoyment was salient in both conditions. To further interpret these results, the contextual use of frequent words was examined using KH Coder's co-occurrence and concordance functions. For the experimental group, *fun/enjoyable* was strongly associated with words such as *learning*, *think*, *this time*, *case*, and *game*. Many participants explicitly compared the VR learning experience to playing a video game. For example, one participant noted, "It was fun because it visually looked like a video game" (translated from Japanese by the author). In contrast, for the control group, *fun/enjoyable* was primarily associated with general terms such as *English* and *learn*. Participants often described the learning experience as enjoyable without specifying the source of that enjoyment. Notably, although the word *fun/enjoyable* appeared eight times in the control group's responses, five of these instances lacked any elaboration on what made the activity enjoyable.

Taken together, while the Likert-scale questionnaire and Mann-Whitney U test indicated a statistically significant difference in enjoyment favoring the VR-TBLT group, the KH Coder analysis suggests that participants in both groups perceived their learning experiences as enjoyable. However, the qualitative patterns indicate differences in how enjoyment was conceptualized and articulated: the VR-TBLT group tended to link enjoyment to game-like elements, whereas the control group expressed enjoyment in more general and less elaborated terms. This discrepancy motivated a more in-depth qualitative exploration of the sources of enjoyment in the follow-up interview analysis.

Detailed Description of Learners' Perceptions

This section reports findings from delayed follow-up interviews, providing a more nuanced account of learners' perceptions of their learning experiences. The interviews focused on two aims: (1) elaborating on factors that made learning enjoyable, and (2) identifying elements that supported long-term retention.

Interview data were analyzed using in vivo coding, which prioritizes participants' own wording to minimize researcher bias (Saldaña, 2016). Codes with similar meanings were subsequently grouped into broader categories to generate overarching themes. For transparency and traceability, the complete set of raw data, in vivo codes, and corresponding categories is provided in Appendices G and H.

What Makes Learning Enjoyable

Table 7 summarizes source of enjoyment for each group based on in vivo codes extracted from participants' open-ended interview responses. The qualitative analysis revealed clear differences in how enjoyment was conceptualized across the two groups. Control-group participants primarily attributed enjoyment to core TBLT characteristics, such as meaning-oriented interaction, authenticity, and practical language use, as well as to familiarity with English learning and the appealing task topic. In contrast, participants in the VR-TBLT group more frequently linked enjoyment to VR-specific affordances—interactivity, immersion, and a sense of presence—while still acknowledging the value of communicative, meaning-focused tasks.

For the traditional TBLT group, four themes emerged. First, participants valued opportunities to use English for practical, real-world communication rather than for test-oriented or grammar-focused study. One participant contrasted this experience with prior classroom learning, stating, "It wasn't the kind of English you learn for tests in high school or grammar-focused stuff—it was practical, usable conversation, and that made it fun." Second, peer interaction enhanced enjoyment by making communication meaningful and socially engaging. This interpersonal dimension was reflected in comments such as, "I could see the other person's facial expressions, so it was easy to tell whether what I was saying was getting across, and that made it fun." Third, several learners described positive emotions associated with reconnecting with English after a long period of limited use. Finally, the familiar and visually appealing context of the ice-cream shop increased motivation and enjoyment. Overall, enjoyment in this group stemmed largely from TBLT's communicative orientation and its contrast with learners' prior classroom experiences.

In the VR-TBLT group, five themes characterized enjoyment. First, physically interacting with the virtual environment (e.g., handling money or objects) made language use more engaging and embodied. As one participant explained, "In high school classes, we just read sentences and imagined the situation. But in VR, I could actually pick up money in the air or hand over objects." Second, collaboration in VR supported enjoyable communication and, for some learners, reduced anxiety through avatar-mediated interaction. This was illustrated by a participant who noted, "It was easier to speak there than face-to-face, because the people I saw were avatars." Third, a strong sense of presence enhanced realism and engagement, with learners reporting a feeling of "being there" in the virtual space. Fourth, immersive and game-like elements transformed learning into a playful experience rather than a purely academic task. Finally, several participants referred to the novelty of VR as a motivating factor, while also acknowledging that this effect might diminish over time, as reflected in comments such as, "Maybe people who usually use VR wouldn't think so, but since it was my first time, that made it enjoyable."

In sum, while enjoyment in traditional TBLT was grounded mainly in communicative authenticity and meaningful interaction, enjoyment in VR-TBLT was driven largely by VR's distinctive affordances. These affordances appeared to amplify TBLT's strengths by making task-based learning more embodied, immersive, and playful, thereby extending its motivational appeal.

Table 7*Source of Enjoyment for both Groups*

	Category	Traditional TBLT	VR-TBLT
TBLT characteristics	Meaning-orientation	✓	✓
	Authenticity	✓	
	Non-linguistic outcomes		
VR Affordances	Interactivity		✓
	Immersion		✓
	Sense of presence		✓
Others	Novelty of technology		✓
	Reconnection with English	✓	
	Topic of target task	✓	

What Makes Learning Memorable

To explore factors contributing to long-term retention, a follow-up interview question was administered to participants who scored above 20 out of 32 on the delayed posttest. Five participants from the VR-TBLT group and one from the traditional TBLT group met this criterion and were asked to reflect on why they were able to recall the target discourse several months later. Although the number of qualifying participants differed across conditions, the responses provide useful insights into perceived retention mechanisms (Table 8).

Table 8*Perceived Factors Supporting Retention of Target Discourse*

	Category	Traditional TBLT (<i>n</i> = 1)	VR-TBLT (<i>n</i> = 5)
Context	Visual-spatial memory	✓	✓
	Episodic-situational memory	✓	✓
	Auditory memory	✓	
Action	Embodied memory		✓

Overall, participants in both conditions reported relying on contextual information—such as scenes, interactions, and experiential cues—rather than isolated linguistic forms when retrieving the target discourse. However, given that a greater number of high-performing learners were found in the VR-TBLT group, VR appears to have supported such contextual recall more consistently.

For the VR-TBLT group, three interrelated themes emerged. First, learners reported that vivid VR scenes functioned as powerful retrieval cues. One participant explained, “By imagining the VR scenes I had seen, I could recall the expressions, and I was able to picture the person I was actually talking to. That made the words come out naturally.” These comments suggest that visual-spatial representations of the virtual environment supported recall.

Second, participants described episodic recall of VR experiences as dynamic events rather than static images. As one learner noted, “Unlike just reading a textbook, studying English in VR felt like an actual experience. Since the interactions, like acting as a customer and clerk..., were realistic, I could remember those situations more easily. I think it also helped me understand how to use the sentences.” This indicates that memory was supported through the replay of situated episodes, closely tied to communicative action.

Third, several participants emphasized the role of embodied action. Physical engagement with the virtual environment appeared to deepen processing and understanding, as illustrated by the comment, “I was able to enter the VR world and use my body while speaking, which helped me focus and think carefully as I learned.” Together, these findings suggest that visual, episodic, and embodied cues jointly contributed to durable memory traces in VR-TBLT.

In contrast, most participants in the traditional TBLT group showed limited retention on the delayed posttest. However, one learner achieved a score comparable to those in the VR group, offering insight into retention mechanisms in a non-VR context. Although based on a single case, her account suggests that successful retention can also occur when language is embedded in rich contextual and interpersonal experiences. She recalled, “I can recall the scenes, the people I did it with, their voices, things like that.” She also described remembering the learning process as a meaningful sequence of activities rather than isolated tasks: “I remember the whole sequence—taking the test, making mistakes, learning the expressions, doing the role-play, and then using them in a way that felt closer to real life.”

Taken together, these findings suggest that contextualized learning—whether supported by VR or by meaning-oriented TBLT activities—plays a central role in long-term retention. While both conditions relied on contextual and experiential cues, VR-TBLT appeared to provide more consistently rich and memorable contexts through visual immersion and embodied action. Given the imbalance in participant numbers, these results should be interpreted cautiously, but they point to VR’s potential for enhancing memory retention through multi-layered contextual encoding.

Discussion

Comparisons between the VR-TBLT and traditional TBLT groups revealed no substantial differences in immediate learning outcomes or transfer performance. Both approaches effectively facilitated target discourse acquisition, listening comprehension, and transfer to a comparable real-world task, indicating that VR-TBLT was at least as effective as traditional TBLT in the short term. This pattern aligns with previous studies reporting mixed or null immediate effects of VR (e.g., Kawasumi & Ishii, 2023; Kobayashi et al., 2023; Satake et al., 2021) and reaffirms the robustness of TBLT as a framework for language development and transfer (Nolen, 2020). Given the small sample size, however, the absence of immediate between-group differences should be interpreted cautiously.

When examining retention under the first research question, the two-month delayed posttest suggestively favored VR-TBLT. Although based on a small, matched subsample and therefore exploratory, this finding aligns with research indicating that VR’s educational benefits may emerge more clearly in delayed or cognitively demanding assessments. For example, Kaplan-Rakowski and Thrasher (2025) reported no immediate advantage for immersive VR but observed marginally significant effects on delayed tests, while Cerezo and Pujolà (2025) similarly concluded that VR benefits may surface over time. Taken together, these findings tentatively suggest that VR-TBLT may support memory consolidation, warranting further investigation with larger samples and longer time frames.

Regarding learners’ perceptions (the second research question), significant differences emerged. Participants in the VR-TBLT group reported higher enjoyment, stronger perceived retention, and greater motivation to continue learning, echoing prior studies that identify enjoyment, confidence, and sustained motivation as key advantages of VR-based language learning (e.g., Kawasumi & Ishii, 2023; Kobayashi et al., 2023; Satake et al., 2021; Smith & McCurrach, 2021). At the same time, the study’s quasi-

experimental design allowed participants to self-select into VR or non-VR conditions, partly to avoid VR sickness. This self-selection may have introduced favorable pre-existing attitudes toward VR, which should be considered when interpreting subjective ratings.

VR-TBLT participants also reported significantly higher perceived difficulty. Interview data suggest that this difficulty stemmed primarily from operational and attentional demands—such as managing movement, object manipulation, and multiple information streams—rather than linguistic complexity. Although participants received brief pre-study training, these findings echo previous reports of cognitive overload in VR environments (e.g., Kaplan-Rakowski & Thrasher, 2025; Yudintseva, 2023) and underscore the need for careful task design, gradual familiarization, and minimization of extraneous cognitive load.

Importantly, higher perceived difficulty did not undermine engagement. Despite finding VR-TBLT more challenging, participants also reported higher enjoyment and motivation, suggesting that the tasks remained manageable and engaging. From a flow perspective, this balance between challenge and engagement may have fostered deeper involvement, potentially contributing to the observed advantage in long-term retention.

Qualitative analyses further clarified group differences. Control-group participants primarily attributed enjoyment to core TBLT features such as meaning-oriented interaction and authentic language use, whereas VR-TBLT participants emphasized VR-specific affordances, including immersion, presence, and interactivity. Although based on a small qualitative sample, these findings suggest that VR may expand sources of enjoyment beyond those typically associated with TBLT. At the same time, participants' references to the "freshness" of VR highlight the possibility of novelty effects (Kim & Namkung, 2024), underscoring the need for longitudinal research to examine whether motivational benefits persist over time.

Qualitative findings also highlighted the role of contextual information in memory retention. High-achieving participants often recalled target discourse by mentally reconstructing learning contexts, consistent with research on context-dependent memory and mental context reinstatement (e.g., Balch et al., 1992; Godden & Baddeley, 1975; Smith, 1979). While this strategy was more frequently reported by VR-TBLT participants—suggesting richer contextual cues in immersive environments—it was also observed in one control-group participant, indicating that contextual encoding is not exclusive to VR.

Finally, VR-TBLT participants frequently attributed retention to embodied, action-based experiences, lending support to theories of embodied cognition in VR language learning (e.g., Cowie & Alizadeh, 2025). Interactions such as handing over objects or navigating space appeared to link language with sensorimotor experience, potentially strengthening memory representations. While embodied learning can occur in physical classrooms, VR's high level of interactivity may offer a distinctive advantage by tightly coupling language, action, and context.

Taken together, the findings offer several practical implications for language teachers considering the adoption of VR-TBLT. First, the results suggest that VR may be a worthwhile investment particularly when instructional goals include long-term retention, contextualized language use, or learner engagement, rather than immediate test score gains alone. Second, both quantitative ratings of difficulty and qualitative interview data highlight the importance of managing cognitive load: sufficient pre-task training, gradual familiarization with VR controls, and task designs that limit unnecessary environmental complexity are crucial for allowing learners to focus on language processing. Third, task design should explicitly exploit VR's unique affordances—such as embodied action, spatialized interaction, and immersive context—by aligning language use with meaningful physical actions and realistic scenarios. Finally, the integration of questionnaire statistics with qualitative analyses of open-ended responses and interviews represents a methodological contribution of this study, demonstrating how mixed-methods approaches can provide a more nuanced understanding of not only whether VR-TBLT works, but how and why learners experience it as engaging, challenging, and memorable. This combination of outcome data and learner perspectives

offers empirically grounded guidance for practitioners seeking to implement VR in pedagogically principled ways.

Limitations and Future Directions

Several limitations should be considered when interpreting the findings of this study. First, the small sample size, particularly for the delayed posttest, limits the generalizability of the results. Because only a subset of participants could be matched across testing phases, the retention findings should be regarded as exploratory. Larger-scale studies are needed to enable more robust statistical analyses and to confirm whether the observed patterns persist across broader learner populations.

Second, the quasi-experimental design, in which participants self-selected into the VR-TBLT or traditional TBLT condition, restricts the strength of causal claims. Although this approach was adopted for ethical and practical reasons (e.g., VR sickness and learner comfort), self-selection may have introduced bias related to learners' prior attitudes toward VR. In addition, slight differences in instructional time and physical engagement between conditions may have influenced outcomes.

Third, the participant pool was limited to young adult learners (ages 18–27), constraining the applicability of the findings to other age groups. Prior research suggests that younger learners may respond differently to VR-based instruction in terms of collaboration and discussion (Rapti et al., 2025). Future research should therefore include a wider age range and examine the feasibility of VR-TBLT in school contexts.

Another limitation concerns potential novelty effects. Participants' motivation and engagement may have been influenced by the newness of VR technology rather than by its pedagogical affordances alone. At the same time, VR use involved physical and cognitive demands, with some learners experiencing fatigue or motion sickness. These findings highlight the need for careful task design, appropriate pacing, and attention to learner well-being.

Future studies should adopt longitudinal or repeated-exposure designs to investigate whether the benefits of VR-TBLT persist over time as novelty effects diminish. Expanding the range of target tasks beyond café interactions to other real-world communicative domains would also help clarify the scope of VR-TBLT's effectiveness. Finally, this study employed VRChat, a customizable social VR platform, rather than a commercial language-learning application. While commercial tools offer convenience, VRChat enabled precise alignment between task demands, environmental design, and target discourse. This suggests that customizable platforms may be particularly valuable for VR-TBLT research and pedagogy when fine-grained control over contextual and interactional features is required.

Conclusion

This study contributes to research on VR-assisted language learning within a Task-Based Language Teaching (TBLT) framework by comparing VR and non-VR instruction under tightly controlled, task-equivalent conditions. While no significant differences were found in immediate language gains or transfer performance, the VR-TBLT group demonstrated significantly better retention of target discourse after a two-month delay, suggesting that VR may support long-term learning even when short-term outcomes are comparable. Qualitative analyses indicate that this advantage is likely related to the rich contextual, episodic, and embodied learning experiences afforded by VR, which helped learners anchor language use to meaningful situations. Methodologically, the study illustrates the value of a mixed-methods approach that integrates performance data with learner perceptions to explain not only whether VR-TBLT is effective, but how and why it may influence learning. Pedagogically, it suggests that while VR may not yield immediate advantages over traditional TBLT, it holds promise for supporting long-term retention and learner engagement when tasks are carefully designed. Theoretically, the findings point to the importance of contextual and embodied factors in second language memory, highlighting VR's potential role in extending TBLT through immersive and action-based learning experiences.

Conflicts of Interest

The author declares no conflicts of interest.

Use of Generative AI

ChatGPT-5 (OpenAI) was used for English editing support; the author is responsible for all remaining errors.

References

- Akyıldız, Y. (2025). The impact of virtual reality on vocabulary learning and cognitive efficiency. *Journal of Educational Technology & Online Learning*, 8(2), 152–163. <https://doi.org/10.31681/jetol.1613357>
- Albayrak, M. S., Chen, J., & Williams, P. J. (2023). A quantitative research on gamifying vocabulary acquisition and retention in virtual reality. *Teaching English with Technology*, 21(2), 42–57. <https://doi.org/10.56297/BKAM1691/DFXC4759>
- Balch, W., Bowman, K., & Mohler, L. (1992). Music-dependent memory in immediate and delayed word recall. *Memory & Cognition*, 20(1), 21–28. <https://doi.org/10.3758/bf03208250>
- Cerezo, L. & Pujolà, J. (2025). 12 digital ludic pedagogies (DLP): Videogames, minigames, extended realities and robots. In J. Muñoz-Basols, M. Fuertes Gutiérrez, & L. Cerezo (Eds.), *Technology-mediated language teaching: From social justice to artificial intelligence* (pp. 270–294). Multilingual Matters. <https://doi.org/10.21832/9781800419889-016>
- Chen, H., & Liao, Y. (2021). Effects of panoramic image virtual reality on the workplace English learning performance of vocational high school students. *Journal of Educational Computing Research*, 59(8), 1601–1622. <https://doi.org/10.1177/0735633121999851>
- Chen, Y., Smith, T., York, C., & Mayall, H. (2019). Google earth virtual reality and expository writing for young English learners from a funds of knowledge perspective. *Computer Assisted Language Learning*, 33(1-2), 1–25. <https://doi.org/10.1080/09588221.2018.1544151>
- Cowie, N., & Alizadeh, M. (2025). Virtual reality for language learning. *ELT Journal*, Article ccaf025. <https://doi.org/10.1093/elt/ccaf025>
- East, M. (2021). *Foundational principles of task-based language teaching*. Routledge. <https://doi.org/10.4324/9781003039709>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66(3), 325–331. <https://doi.org/10.1111/j.2044-8295.1975.tb01468.x>
- Kaplan-Rakowski, R., & Thrasher, T. (2025). The impact of high-immersion virtual reality and interactivity on vocabulary learning. *British Journal of Educational Technology*, 56(6), 2647–2670. <https://doi.org/10.1111/bjet.13603>
- Kawasumi, S., & Ishii, Y. (2023). Comparative study of English learning using virtual reality and a smartphone application. *Bulletin of the Faculty of Education, Chiba University*, 71, 99–105. <https://opac.ll.chiba-u.jp/da/curator/900121724/>

- Khezlrou, S. (2025). Task-based language teaching. In Y. Wang & G. Stockwell (Eds.), *The Cambridge handbook of technology in language teaching and learning* (pp. 377–392). Cambridge University Press. <https://doi.org/10.1017/9781009294850.028>
- Khodabandeh, F. (2022). 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, Article 27. <https://doi.org/10.1186/s41239-022-00334-w>
- Kim, Y., & Namkung, Y. (2024). Methodological characteristics in technology-mediated task-based language teaching research: Current practices and future directions. *Annual Review of Applied Linguistics*. Advance online publication. <https://doi.org/10.1017/S0267190524000096>
- Kobayashi, W., Judai, K., Ozaki, T., Kanda, T., Schmidt, M. G., Kamozaawa, S., & Tani, S. (2023). The advantages and challenges of VR-enhanced foreign language learning in a Japanese university: A preliminary study. *Bulletin of the Institute of Humanities, College of Humanities and Sciences, Nihon University*, 106, 173–196. <https://doi.org/10.15006/chs20101106010>
- Nolen, C. (2020). *Examining task transferability in task-based language teaching: A multi-case study* [Doctoral dissertation, Georgia State University]. ScholarWorks @ Georgia State University. https://scholarworks.gsu.edu/alesl_diss/53
- Rapti, S., Sapounidis, T., & Tselegkaridis, S. (2025). Investigating educators' and students' perspectives on virtual reality enhanced teaching in preschool. *Early Childhood Education Journal*, 53, 1107–1118. <https://doi.org/10.1007/s10643-024-01659-z>
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage.
- Satake, Y., Yamamoto, S., & Obari, H. (2021). Effects of virtual reality use on Japanese English learners' foreign language anxiety. *ICERI2021 Proceedings*, 1234–1240. <https://doi.org/10.21125/iceri.2021.0358>
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, 5(5), 460–471. <https://psycnet.apa.org/doi/10.1037/0278-7393.5.5.460>
- Smith, M., & McCurrach, D. (2021). The usage of virtual reality in task-based language teaching. *Proceedings of the 28th Korea TESOL International Conference: Re-envisioning ELT Altogether, All Together, KOTESOL*, 28(1), 153–165. <https://koreatesol.org/content/kotesolproceedings>
- Taguchi, N., & Zhao, S. (2025). Can virtual reality leverage technology-mediated task-based language teaching? A research synthesis. *Digital Studies in Language and Literature*, 2(1), 25–53. <https://doi.org/10.1515/dsll-2024-0028>
- Xie, Y., Chen, Y., & Ryder, L. (2019). Effects of using mobile-based virtual reality on Chinese L2 students' oral proficiency. *Computer Assisted Language Learning*, 34(3), 225–245. <https://doi.org/10.1080/09588221.2019.1604551>
- Yan, W., & Lowell, V. L. (2024). Design and evaluation of task-based role-play speaking activities in a VR environment for authentic learning: A design-based research approach. *The Journal of Applied Instructional Design*, 13(4), 309–327. <https://doi.org/10.59668/2033.18647>
- Yudintseva, A. (2023). Virtual reality affordances for oral communication in English as a second language classroom: A literature review. *Computers & Education: X Reality*, 2, Article 100018. <https://doi.org/10.1016/j.cexr.2023.100018>

Appendices

Find appendices in the [Supplementary Materials](#) online.

About the Author

Koyo Ogawa is a doctoral student at the Prefectural University of Kumamoto and an incoming Assistant Professor at Nagasaki International University. His research explores how VR-enhanced task-based language teaching can facilitate learners' transfer of second language skills to real-world contexts.

E-mail: ogako.1124@gmail.com

ORCID: <https://orcid.org/0009-0003-6407-4172>