

論文 / 著書情報
Article / Book Information

Title	Virtual reality study abroad and language contact patterns in Japanese EFL learners
Authors	Olexa Robert Anthony, Taquet David
Citation	Discover Education, Volume 5, ,
Pub. date	2026, 1
DOI	https://dx.doi.org/10.1007/s44217-026-01144-2
Creative Commons	Information is in the article.

RESEARCH

Open Access



Virtual reality study abroad and language contact patterns in Japanese EFL learners

Robert Anthony Olexa^{1,2*} and David Taquet¹

*Correspondence:

Robert Anthony Olexa
olexa@hakodate-ct.ac.jp
¹General Education Department,
National Institute of Technology,
Hakodate College, Hakodate,
Hokkaido, Japan
²Global Engineering for
Development, Environment
and Society, Tokyo Institute of
Technology, Tokyo, Japan

Abstract

Japanese English learners in virtual reality (VR) experienced an unexpected reversal: they served as linguistic resources for others four times more often than did those in physical study abroad. This role reversal suggests that VR environments may create fundamentally different interactional dynamics than their physical counterparts. While study abroad programs effectively develop English proficiency, high costs and institutional barriers limit access to these programs. VR offers an alternative, yet comparative data on language exposure in virtual and in-person exchanges remain scarce. This exploratory mixed-methods study ($N = 14$) provides the first application of the Language Contact Profile (LCP) to VR environments, comparing English exposure across three formats: a six-week Belgian exchange program, a three-week program in Singapore, and a synchronous VR experience ranging from one to four weeks. The results show that VR functions as a mediated 'third space' that redistributes rather than reduces language contact. Japanese participants served as L1 resources significantly more often in VR than in Belgium ($\epsilon^2 = 0.65$), suggesting role reversal where learners became teachers. Compared with those in physical programs, VR participants reported less instructor contact ($\epsilon^2 = 0.68$) and fewer incidental overhearing opportunities ($\epsilon^2 = 0.78$), whereas Singapore yielded the highest level of peer interaction. Despite these constraints, VR participants independently initiated service encounters and compensated through self-directed learning. We emphasize effect sizes and robust statistics (means/medians/IQRs). VR complements physical programs by offering distinct interactional dynamics. This paper provides practical implementation guidance while acknowledging the need for larger multisite validations.

Keywords Virtual reality, Study abroad, Mixed methods, Language contact profile, English as a foreign language

1 Introduction

In Japan, English as a foreign language (EFL) exposure rarely extends beyond the classroom. A typical class of 40 students meets for 90 min each week, with ample time for grammar and vocabulary drills but not for sustained discourse or cultural context. While study abroad programs conventionally fill this gap, participation has declined in Japan [1], with costs and institutional barriers limiting access.



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

In search of broader access, universities have turned to online formats such as collaborative online international learning (COIL) and even smartphone-mediated exchanges [2, 3]. Virtual reality (VR) in particular promises something new: a richly embodied environment where learners converse with global English speakers as if they were physically co-present. However, this promise remains untested: no systematic study has compared language exposure in VR to that in traditional study abroad via a common metric.

This study addresses this gap by applying the Language Contact Profile (LCP) developed by Freed, Dewey, and Segalowitz [4] to compare three exchange formats: Belgium (6 weeks), Singapore (3 weeks), and a synchronous VR experience (1–4 weeks). Fourteen Japanese engineering students completed pre- and post-LCP surveys, as well as weekly language logs.

We asked:

- (1) How does VR redistribute language contact opportunities compared with physical study abroad?
- (2) What are the differences in frequency, context, and exposure among EFL students in Belgium, Singapore, and VR?

This study makes three contributions. First, we provide the first application of the Language Contact Profile (LCP) to synchronous immersive VR, establishing a baseline for comparing virtual and physical immersion. While Davidson and Garas [5] tracked time-on-task in emergency remote programs, they did not employ standardized contact profiling or specifically examine VR. Our application thus provides a first comparable baseline with decades of study abroad research while identifying where the instrument requires adaptation for virtual contexts (e.g., distinguishing scripted NPCs from spontaneous peer exchanges). Second, we document an unexpected role reversal, where Japanese EFL learners serve as L1 resources for nonnative Japanese speakers in VR significantly more than they do in physical programs ($\eta^2 = 0.65$, $p = 0.015$), challenging traditional novice–expert dynamics. Third, we theorize VR not as a diminished form of physical immersion but as a qualitatively different hybrid space with unique pedagogical affordances.

The literature review examines the benefits and limitations of studying abroad and then explores how VR might restructure rather than replicate these experiences.

2 Literature review

2.1 History of study abroad

The European Grand Tour, in which aristocrats from the 16th to the 19th centuries completed their education through international travel, established study abroad as a transformative educational experience [6]. This elite tradition evolved as universities institutionalized study abroad programs between world wars, expanding participation [7, 8]. In Japan, study abroad programs emerged during the Meiji period as part of the country's modernization efforts. Early Japanese study abroad programs initially focused on sending women abroad to learn Western customs and etiquette, later shifting toward academic achievement. Over time, more participants returned with advanced degrees, which led to the establishment of university programs informed by their experiences abroad [8]. This trajectory, from elite privilege through institutional programs to accessibility challenges, continues with the emergence of today's virtual alternatives.

Today, study abroad programs are commonly offered in higher education institutions worldwide. Participation in study abroad programs worldwide remains relatively small but has been growing steadily. In contrast, Japan has recently experienced a decline in student participation in study abroad programs [1]. Most study abroad research examines contexts where English functions as either the target language or lingua franca [9], which is the category relevant to our Japanese participants studying in Belgium, Singapore, and VR.

2.2 Benefits of study abroad

Meta-analyses reveal a medium-to-large overall advantage for studying abroad over domestic instruction (Hedges' $g = 0.87$) [10], with robust gains in oral fluency and sociolinguistic competence, although grammatical accuracy shows minimal differences. These differential gains reflect environmental affordances: classroom learning develops declarative knowledge, whereas immersion fosters procedural fluency [11].

However, critics argue that these benefits have been idealized, noting inconsistent program goals and mixed empirical evidence [12]. Program variables (duration, cultural context, and residence type) complicate outcome prediction. Llanes et al. [13] reported that even in English-dominant contexts, proficiency gains vary unpredictably, with pre-test scores failing to predict improvement, contradicting DeKeyser's [11] findings—individual factors (motivation, social networks, interaction quality) further moderate outcomes. A study of Japanese students in the UK revealed that improvements in pragmatic competence (or lack thereof) were linked to these interrelated factors [14].

Given mixed but generally positive study abroad outcomes, educators are testing virtual alternatives.

2.3 Virtual alternatives: opportunities and constraints

Virtual exchanges, facilitated through video conferencing, gaming, and VR, offer both synchronous and asynchronous practice opportunities, thereby increasing access while reducing costs. VR uniquely enables multimodal interaction—gesture, gaze, and object manipulation—supporting embodied, layered ('laminated') meaning-making [15], where communication builds meaning iteratively. We adopt the term virtual reality study abroad (VRSA) to distinguish immersive VR from other virtual formats, corresponding to the Japanese term "*virtual ryūgaku*" (バーチャル留学).

Physical study abroad embeds learners in the accumulated meanings of places. Consider, for example, the löyly, water hitting the heated rocks filling a Turku neighborhood sauna with steam, which carries centuries of Finnish social practice, or the cicadas harmonizing outside a Tokyo dormitory window, marking the rhythm of Japanese summer. These sensory anchors permeate routine interaction and the wider discourse of poetry and prose. Reproducing such context-rich cues in VR is challenging: place-bound signals moor language to lived experience and are hard to simulate with fidelity.

Because physical immersion braids language to place, VR seldom reproduces that thick texture. Paradoxically, those absences flip into affordances: unbound by geography or travel costs, VRSA learners meet deterritorialized interlocutors in brief, self-selected exchanges where linguistic roles are fluid—Japanese learners may momentarily become teachers. Technological mediation (including latency, audio mixing, and avatar proxemics) reshapes turn-taking [16]. These constraints enable safety (easy exit),

partner choice, and designed tasks (scripted NPC service encounters, on-demand role-play, ambient listening zones). In our data, contact is redistributed and not reduced—VR complements rather than imitates physical immersion.

Understanding these distinctions requires theoretical grounding. We conceptualize VR as a ‘mediated third space’—a hybrid context with unique pedagogical possibilities.

3 Theoretical framework

Building on sociocultural and situated learning perspectives, we frame immersive-VR language exchanges as communication systems in which learners co-construct knowledge through context-embedded interactions. Within this lens, VR constitutes a qualitatively different study abroad environment for EFL learners.

3.1 Virtual reality as a mediated third space

We conceptualize platforms such as Bigscreen Beta and VRChat as ‘mediated third spaces’, hybrid environments where features of home and target cultures coexist [17]. We operationalize this with three observable criteria: participants are (a) physically in Japan, (b) socially co-present in the same virtual world, and (c) mobilize both L1 and L2 resources to co-construct meaning.

A central reason VR is *not* merely a simulated space is bilocation: learners remain bodily anchored while mentally constructing an elsewhere. Evidence from hippocampal place-cell research shows that spatial coding in VR is attenuated or altered relative to physical navigation (e.g., weaker activation in humans; mixed findings in rodents), indicating a hybrid spatial-memory mode rather than full spatial equivalence to the real world [18–20]. In this view, VR-language learning spaces can function like a memory palace: spatially organized yet transparently artificial environments that leverage spatial cognition for encoding and retrieval of cultural-linguistic knowledge. VR’s sensory gaps (absence of heat, taste, and reduced haptic feedback) continuously signal its own artificiality. Bilocation is pedagogically productive because it is incomplete.

3.2 Linguistic marketplace

Bourdieu’s [21] linguistic marketplace explains how social settings assign value to linguistic resources—and to the speakers that hold them—depending on who is present and what counts as legitimate language use.

In-person programs (Belgium, Singapore) are particularly beneficial because interlocutor pools are geographically bound and institutionally structured, which lends English a stable value. In Belgium, students reported using English more frequently with service personnel and teachers than with their classmates, which is consistent with role expectations in those relationships. In Singapore, English’s official status facilitated broader use; local varieties (e.g., Singlish) likely increased ambient overhearing opportunities. In both cases, the market tilts toward consuming English, with few occasions for Japanese learners to act as language providers.

In VR communities such as EN-JP [22], an organic English/Japanese language exchange on the VR communication platform VRChat [23], community composition can shift the market. When many regulars seek Japanese, the value of Japanese increases relative to English, inviting role-reversal episodes in which Japanese learners temporarily occupy expert roles even while pursuing L2 English goals. This marketplace tilt

is predictable, and our results show higher frequencies of Japanese as a resource in VR than in Belgium.

3.3 Iconicity as pedagogical design

Rather than pursuing photorealistic simulation, we approach VR as a semiotic medium whose iconic resources can be orchestrated for language learning. VR affords multiple iconicities—visual-spatial (scale, proximity), auditory (spatialized sound), and limited haptic—that pair with linguistic iconicity to stabilize form-meaning mappings [24]–[25]. Some deployments intentionally exaggerate scales or motion to make critical relations perceptible, privileging cue placement over fidelity, which is consistent with VR training scenarios that prioritize scale-specific experiences [26]. Learning gains depend on how iconic cues scaffold interaction at critical moments, not on realism.

3.4 Redistributed interaction opportunities

Kern, Liddicoat, and Zarate identify seven interaction factors (ranging from technical interfaces to intersubjectivity) that manifest differently in VR contexts than in physical contexts [17]. Avatar design mediates identity; audio latency and mixing affect turn-taking; gesture libraries constrain expression. These technological mediations redistribute rather than reduce opportunities. When iconic cues scaffold meaning-making, Japanese learners unexpectedly become L1 resources, suggesting that VR creates qualitatively distinct pedagogical approaches.

To test these propositions, we compare language contact in in-person studies abroad and VRSA using a mixed-methods design, linking LCP responses with session-level evidence (e.g., logs) to evaluate how designed affordances map to observed interaction patterns.

4 Research and design methods

This mixed-methods study compared language contact across three institutional programs: Belgium ($n=6$, 6 weeks), Singapore ($n=3$, 3 weeks), and VR study abroad (VRSA) ($n=5$, 1–4 weeks). Duration differences reflect institutional constraints; we address comparability through the use of effect sizes and normalized frequency measures.

4.1 Participants

After institutional ethics approval was obtained, we recruited 14 volunteer engineering students (6 females, 8 males; median age, 19) from a Japanese technical college. Thirteen were Japanese nationals; one was Malaysian. The median pre-program TOEIC score was 527 ($\approx$ B1 CEFR), indicating lower-intermediate proficiency. Given the small sample size, we emphasize effect sizes and robust statistics over p values (see Analysis).

4.2 Program implementations

Physical programs in Belgium and Singapore offered English–medium research internships without formal language courses. Belgium students lived in shared apartments (5) or with a host family (1); Singapore students lived in university residences. Both generate routine service encounters and frequent contact with instructors.

Virtual Reality Study Abroad (VRSA) Participants borrowed Meta Quest HMDs for home use following safety orientation (30-minute sessions, motion sickness screening).

They primarily used VRChat EN-JP [22] and Bigscreen [27], with optional applications such as Mondly [28], for role-playing scenarios, including hotel check-ins and taxi rides. Unlike webcam exchanges, 6DoF embodiment enables gestures and proxemics; open-world design exposes learners to heterogeneous interlocutors (varying in age, proficiency, and cultural background), approximating the diversity of in-person encounters (see Fig. 1).

4.3 Measures

Language contact profile The LCP quantifies target-language exposure through direct interaction (instructors, peers, and service personnel) and indirect channels (overhearing and media) [4, 29]. The framework provides a systematic approach to understanding the extent and types of target-language practice that different learners actually engage in outside of class.

We retained original LCP items for comparability with existing study abroad, despite their development for physical contexts. Some items required interpretive flexibility—“talking with service personnel” encompassed both scripted role-play and live interactions; “overhearing” included ambient virtual audio. A full discussion of measurement limitations can be found in the Limitations section.

Language Logs: For course credit, all the students submitted Japanese-language logs with four prompts: (1) internship activities, (2) challenges or noteworthy observations, (3) goals for the next week, and (4) general reflections on life abroad or in VR. Although some entries were brief, limiting thematic depth, the logs added valuable context.

4.4 Analysis

Between-group differences were tested via the Kruskal–Wallis test (χ^2 , p , ϵ^2) with Dwass–Steel–Critchlow–Fligner (DSCF) post hoc comparisons. The DSCF procedure controls familywise error across all pairwise comparisons; no additional Holm correction was applied. Analyses were conducted in jamovi 2.6 [29], which utilizes R (Version 4.4) [38] and the car package [39]. Rank-biserial effect sizes were calculated using Kerby's [40] simple difference formula. (see Table 1).



Fig. 1 Students Interacting in the Bigscreen App. Left: Student wearing the headset during an activity. Right: In-app view of avatars engaged in conversation

Table 1 Language contact profile: significant differences in self-reported english exposure across three study abroad modalities

Language Use Context	Belgium (n = 6) $\bar{x}$ /Mdn/IQR	Singapore (n = 3) $\bar{x}$ /Mdn/IQR	VR (n = 5) $\bar{x}$ /Mdn/IQR	χ^2	p value	ϵ^2
Talking with the instructor	4.3/5.0 [5.0–5.0]	4.7/5.0 [4.5–5.0]	1.0/1.0 [1.0–1.0]	8.89	0.012*	0.68
Talking with classmates	2.8/1.0 [1.0–4.8]	5.0/5.0 [4.0–6.0]	1.0/1.0 [1.0–1.0]	6.17	0.046*	0.48
Talking with service personnel	3.3/4.0 [1.7–4.8]	5.7/7.0 [5.0–7.0]	1.6/1.0 [1.0–1.0]	5.50	0.064	0.42
Speaking Japanese to nonnative speakers	1.0/1.0 [1.0–1.0]	3.7/4.0 [3.0–4.5]	4.0/4.0 [3.0–5.0]	8.44	0.015*	0.65
Overhearing English conversations	1.0/1.0 [1.0–1.0]	4.7/4.0 [3.5–5.5]	1.4/1.0 [1.0–1.0]	10.12	0.006**	0.78

The values represent days per week (0–7)

* $p < 0.05$, ** $p < 0.01$

5 Results

Group baselines were statistically equivalent. Four of the 32 LCP items differed across groups with large effect sizes ($\epsilon^2 > 0.40$): instructor contact ($\epsilon^2 = 0.68$), role reversal (Japanese as a resource) ($\epsilon^2 = 0.65$), overhearing opportunities ($\epsilon^2 = 0.78$), and talking with classmates ($\epsilon^2 = 0.48$), suggesting meaningful program characteristics rather than sampling variability. For significant omnibus tests, we conducted DSCF post hoc comparisons with rank-biserial r values for effect sizes (small ≈ 0.10 , medium ≈ 0.30 , ≥ 0.50). Complete pairwise comparisons appear in Appendix Table 3 and 4.

5.1 Role reversal: Japanese learners as linguistic resources

The number of days per week speaking Japanese to nonnative Japanese speakers differed significantly across the groups, $\chi^2(2) = 8.44$, $p = 0.015$, $\epsilon^2 = 0.65$ (large). Post hoc comparisons revealed that students in the VR group reported significantly more frequent interactions with nonnative speakers than did those in Belgium ($p = 0.030$, $r = 0.80$). Singapore participants also reported more frequent interactions than Belgium did ($p = 0.016$, $r = 1.00$). There were no significant differences between Singapore and VR ($p = 0.988$). On the basis of the medians, the participants reported speaking Japanese to nonnative speakers four times as often on the VRSA as they did in Belgium (4.0 vs. 1.0 days/week).

5.2 Instructor contact

The Kruskal–Wallis test revealed a significant group effect, $\chi^2(2) = 8.89$, $p = 0.012$, $\epsilon^2 = 0.68$ (large). Post hoc DSCF comparisons revealed that VR participants had significantly less instructor contact than both Singapore ($p = 0.027$, $r = -1.00$) and Belgium ($p = 0.023$, $r = -0.83$) did. The rank-biserial correlation of $r = -1.00$ between Singapore and VR indicates a complete separation between the groups, with no overlap in instructor contact frequency scores. The results reflect the structured nature of Singapore's program (daily instructor supervision) versus VR's self-directed approach (minimal formal instruction). VR participants averaged ~ 1 day per week of instructor interaction versus 4.3 (Belgium) and 4.7 (Singapore).

5.3 Talking with classmates

Students' reports of how frequently they spoke with classmates differed significantly across the programs ($\chi^2(2) = 6.17$, $p = 0.046$, $\epsilon^2 = 0.48$). Post hoc comparisons revealed

that Singapore participants interacted with their classmates more frequently than the VR participants did ($p = 0.028$, $r = -1.00$). Belgium was not significantly different.

5.4 Talking with service personnel

Interactions with service personnel typically occur outside institutional settings (e.g., giving directions to a taxi driver or ordering food). No between-group differences were detected; DSCF pairwise comparisons were not significant (all $p > 0.05$). VR participants reported one English interaction with ‘service personnel’ per week. Given mixed sources (scripted NPCs vs. live transactions), this likely reflects structured service encounter practices rather than naturally occurring in-world services.

5.5 Overhearing/ambient listening

A significant group effect was found for students’ attempts to overhear English conversations outside of class, $\chi^2(2) = 10.12$, $p = 0.006$, $\varepsilon^2 = 0.78$ (large). Post hoc tests revealed that Singapore had more attempts to overhear than Belgium did ($p = 0.016$, $r = 1.00$). Comparisons with VR were not significant (Singapore vs. VR: $p = 0.065$; Belgium vs. VR: $p = 0.517$). The results reflect structural differences in ambient language exposure (see Table 1).

5.6 Language logs

Thematic analysis of weekly language logs yielded seven themes (Table 2). We highlight representative evidence and link RQ1 (redistribution of contact) and RQ2 (frequency/contact differences). Quotes are presented as originally written in English and translated from Japanese; all are anonymized.

Table 2 Codes and themes identified in Language logs

Themes	Subthemes (n excerpts)	Example quotes
1. Communication challenges	Difficulty with complex topics (4) Urgency and spontaneity (1) Code-switching (2) Miscommunications (3)	<i>I found that I'm good at discussing daily things, but I'm not good at having important conversations.</i> [BEL]
2. Cultural and contextual influences	Cultural misunderstanding (2) Cultural learning (4)	<i>I tried to tell my house owner about fireworks in Japan, but I couldn't tell her well.</i> [BEL]
3. Learning strategies	Vocabulary acquisition (2) Feedback and imitation (2)	<i>My pronunciation got better — someone told me that because I mimicked how native speakers said it.</i> [VR]
4. Social and emotional aspects	Relationship building (2) Emotional discomfort (2) Encouragement and compliments (2)	<i>A friend became willing to take photos with me.</i> [VR]
5. Technology and virtual environments	Safe space in VR (1) Multilingual VR interaction (3) VR communication challenges (1)	<i>I think being able to disconnect quickly when I feel uncomfortable is one good thing about VRChat.</i> [VR]
6. Motivation and personal growth	Motivation from small wins (2) Gratitude and engagement (1)	<i>When I said, “Dank je well,” my house owner complimented me. I decided to study Dutch.</i> [BEL]
7. Practical and transactional use	Food and shopping (4) Information-seeking (2) Contextual difficulty (1)	<i>I tried to order an ice cream, onigiri, and mochi ice.</i> [BEL]

Numbers in parentheses indicate the frequency of participant references to each subtheme

- (1) **Communication Challenges (RQ1)** Students struggled with complex meanings and cultural nuances in both in-person and VR contexts:

I tried to tell in English that the meeting with my friend didn't go well, and I couldn't get in touch with her, but it didn't communicate too well. [SIN]

- (2) **Cultural and Contextual Influences (RQ2)** Students reported occasional contextual and cultural misunderstandings that shaped how the messages were interpreted:

The stores around I live can't use credit cards to pay, so I confused about buy something. [BEL]

- (3) **Learning Strategies (RQ1)** The learners reported immediate repair and lookup:

Every time I heard new words, I always ask the meaning or search it right after the conversation. [VR]

- (4) **Social-emotional Aspects (RQ1/RQ2)** The emotional dimensions of language learning were closely linked to confidence, feedback, and social connection:

I had a friend take me to the anime festival. It was so fun. [SIN]

However, some interactions led to discomfort or emotional fatigue for participants:

I have a lot of scary dreams in the night. [BEL]

These emotional fluctuations demonstrate how real and virtual interactions can contribute to personal growth or expose new forms of vulnerability.

- (5) **Technology and virtual environments (RQ1)** VR offers multilingual contact and safety features but also introduces challenges such as identity ambiguity.

(A VRChat user) started saying romantic things, and someone I thought was a woman turned out to be a man using a voice changer. It was stressful. [VR, translated]

- (6) **Motivation and personal growth (RQ2)** Positive reinforcement from interlocutors and their encouragement contributed significantly to self-esteem and motivation:

When I said, "Dank je wel," my house owner complimented me. I decided to study Dutch. [BEL]

- (7) **Practical and transactional use (RQ2)** Much of the participants' language use was practical, involving ordering food, navigating daily needs, or asking questions about events or systems.

I wanna know whether I can use credit card to pay in Brussels international airport. [BEL]

6 Discussion

Language contact patterns differed significantly across the three study abroad modalities. The analysis revealed that VR functions as a mediated third space, redistributing rather than reducing opportunities for language contact. Most notably, Japanese participants served as L1 resources significantly more often in VR than in Belgium ($\epsilon^2 = 0.65$) while reporting less instructor contact ($\epsilon^2 = 0.68$) and fewer incidental overhearing opportunities ($\epsilon^2 = 0.78$). The Singapore participants presented the highest peer

interaction rates. Despite structural constraints, VR participants independently initiated service encounters and compensated through self-directed learning strategies. These patterns suggest that VR complements rather than replicates physical study abroad, offering distinct interactional dynamics with implications for program design.

6.1 Convergence of quantitative and qualitative findings

Triangulating LCP data with language logs revealed consistent patterns in three key areas central to RQ1/RQ2.

Role Reversal: Mechanism and frequency The role reversal phenomenon, in which VR and Singapore participants served as Japanese language resources more frequently than Belgium participants did ($\epsilon^2 = 0.65$), emerged consistently across both the quantitative and qualitative data. The LCP documented the frequency—VR participants reported speaking Japanese to non-native speakers four times as often as Belgium participants did (median 4.0 vs. 1.0 days/week)—while language logs illuminated the underlying mechanism.

VR's global user base included non-native Japanese speakers who, upon encountering Japanese students, initiated Japanese practice. One participant described this dynamic:

I always talk in English, but when they don't understand, I translate it in Japanese.

Another reported:

On VRChat, there was a Vietnamese person who told me that he would accompany me to practice English every day, so I asked him to be my practice partner.

These exchanges exemplify how VR communication, which is freed from geographic constraints but bounded by access to technology, enables encounters with linguistically diverse interlocutors—Vietnamese English learners and aspiring Japanese speakers from various L1 backgrounds—unavailable in geographically bounded programs where interlocutor pools reflect local demographics.

The pattern aligns with Bourdieu's linguistic marketplace framework: in physical programs, institutional structures, and geographic constraints, English was positioned as the dominant currency, with Japanese students functioning primarily as language consumers. In VR's EN-JP world (a bilingual community where participants explicitly seek both English and Japanese practice), the composition of the marketplace shifts. Japanese culture holds increased value for interlocutors seeking Japanese practice, enabling role reversal episodes in which Japanese students temporarily occupy expert positions. Students were aware of EN-JP's bilingual focus when selecting it, suggesting that they anticipated or sought contexts where their L1 might function as a resource rather than viewing themselves solely as L2 learners.

Impact of the interaction structure Compared with VR participants, Singapore participants presented significantly greater institutionally structured peer contact—talking with classmates 5.0 days/week ($\epsilon^2 = 0.48$)—reflected in logs describing repeated encounters, where students wrote about visiting an anime festival and taking photos with friends.

The spatial-temporal settings of physical programs (dormitories, classrooms, cafeterias) create persistent social contexts where repeated interaction with peers enables

relationship development and stabilizes interactional roles. Students became known to their classmates, establishing predictable linguistic hierarchies.

The ephemeral encounter structure of the VR contrasts sharply. Students met diverse interlocutors across sessions but rarely developed sustained relationships within VR itself, instead compensating through external platforms:

I use the World of Japanese and English speakers [22]. I could speak to people around the world and make friends. I exchanged my Discord handles with one such person. VR is very useful.

The shift from institutional to self-generated interaction contexts changed the quality of linguistic practice. One student reflected:

Earlier, I had never had the experience of speaking English with someone other than my teacher for a long time, so I could not only learn English conversation but also how to communicate with others.

When classroom interaction follows predictable teacher-fronted patterns, VR peer exchanges require authentic communicative breakdowns, identity negotiations, and topic shifts without institutional safety nets. Students cannot rely on teachers to repair misunderstandings or simplify input; they become responsible for maintaining interaction through strategic code-switching, clarification requests, and repetition—negotiation processes that generate comprehensible input and facilitate acquisition [34].

The recurring unfamiliarity and encounters with unexpected interlocutors created conditions in which linguistic roles remained negotiable. Whether a Japanese student functioned as an L2 learner or an L1 resource was determined anew in each encounter, potentially explaining why role reversal occurred as frequently in VR as in Singapore (median 4.0 vs. 3.7 days/week, $p=0.988$), despite dramatically different peer contact structures.

Although role reversal occurred significantly more in VR than in Belgium, students rarely framed these episodes as pedagogically noteworthy. The logs contained few explicit reflections on “teaching Japanese” or “switching roles,” suggesting that students experienced these exchanges as routine communication repairs rather than identity transformations. More critically, the logs provided limited evidence that role reversal episodes involved intercultural practices, such as negotiating cultural differences, explaining culturally embedded concepts, or navigating cross-cultural misunderstandings. As Liddicoat [17] asserts, “a much more satisfying way of framing the question is to ask ‘to what extent do participants in such interactions employ intercultural practices?’” When framed in this way, the answer requires empirical observation of interaction specifics rather than relying solely on frequency counts.

We cannot distinguish whether (a) students genuinely experienced role reversal as unremarkable functional code-switching within VR’s multilingual ecology, (b) role reversals involved shallow linguistic exchange without deeper intercultural engagement, or (c) our weekly prompts failed to elicit relevant reflections about cultural negotiation. Future research employing conversation analysis of recorded VR sessions could illuminate whether Japanese-as-resource episodes involve merely lexical translation or whether they create opportunities for metalinguistic and intercultural awareness, such

as explaining cultural references, discussing pragmatic differences, or reflecting on L1/L2 structural contrasts.

6.2 Self-directed learning and autonomous compensation

The quantitative finding of reduced instructor contact in VR ($M \approx 1$ day/week versus 4.3 days/week in Belgium, 4.7 days/week in Singapore; $\varepsilon^2 = 0.68$) aligned with qualitative themes of self-directed learning and linguistic autonomy (Theme 3: Learning Strategies). Following Seliger's [30] framework, which distinguishes between high- and low-input-generating learners, VR participants function as active interaction seekers in unstructured environments outside institutional scaffolding. VR participants explicitly noted compensatory strategies (immediate dictionary lookups, meaning clarification requests) and reliance on peer feedback for pronunciation improvement rather than instructor correction (see Table 2, Theme 3).

Physical programs embed students in institutional frameworks with designated instructors, scheduled check-ins, and supervised activities. VR participants lacked such scaffolding but exercised greater control over interaction timing, partner selection, session duration, and topic choice. The trade-off appears in Theme 4 (social-emotional aspects): some students valued VR's safety affordances, noting the ability to disconnect immediately when they felt uncomfortable. In contrast, others in physical programs reported stress without equivalent escape mechanisms (see Table 2, Theme 4).

This pattern may reflect participant selection rather than modality effects. VR participation was voluntary and self-directed, potentially attracting students who were already inclined toward autonomous learning and high-input generation. Language logs revealed that even self-selected VR participants varied in engagement levels, with some expressing difficulty maintaining practice due to competing academic demands. As one student wrote,

My excuse is that I didn't do enough VR activities due to assignments and other commitments. Sorry.

Reduced instructor contact represents a program design choice rather than VR necessity. A VR study abroad program could integrate regular instructor check-ins via video conferencing or in-world meetings; our implementation prioritized autonomous exploration. The self-directed approach aligns with *virtual ryūgaku* practices, emphasizing horizontal peer learning over vertical teacher-led instruction [31]. Future VRSA implementations might deliberately vary instructor involvement to isolate modality effects from pedagogical approach effects while controlling for participant self-selection through random assignment or mandatory participation requirements.

6.3 Synthesis

The convergence validates our mixed-methods approach. The quantitative LCP data identified what differed across contexts (contact frequency by interlocutor type). Qualitative logs explained how and why these differences manifested (institutional structure versus autonomous seeking, relationship continuity versus ephemeral encounters, instructor scaffolding versus peer/self-directed repair). The datasets suggest redistribution, not reduction: self-initiated, peer-mediated, role-flexible interactions increase,

whereas instructor-led, institutionally structured, and hierarchically stable interactions recede.

The brevity of weekly entries limits phenomenological depth. We did not code specifically for role reversal episodes, bilocation experiences, or platform-specific affordances despite foregrounding these theoretically. Photoelicitation interviews, social media trace analysis, or stimulated recall protocols using screen recordings [32] could capture in-the-moment awareness unavailable through weekly retrospection, providing richer data to test the theoretical frameworks we apply interpretively in the following section.

6.4 Interpreting VRSA through theoretical lenses

The traditional study abroad research, which is grounded in interactionist [33] and sociocultural theories [34], assumes that learners benefit from encountering language within geographically bounded speech communities with shared linguistic norms. Physical programs immerse students in contexts where language is associated with specific places and populations, which sociolinguistics recognizes as speech communities with shared linguistic norms (e.g., Singlish in Singapore or English as a Lingua Franca in Belgium). Language populations facilitate transparent connections between linguistic forms and the cultural-geographic referents that underpin rationales for immersion and study abroad.

However, as Murgiano et al. [25] argue, language is inherently situated through multimodal iconicity and indexicality—pointing to and resembling real-world referents. VRSA reconfigures these conditions in ways that require theoretical consideration. Sauro & Zourou [35] described the “digital wilds” as autonomous and informal language learning spaces where learners exercise high agency in seeking practice opportunities. Platforms such as VRChat [23] enable such autonomous exploration. Our participants discovered and self-selected into EN-JP [22] on the basis of personal learning goals rather than institutional assignment. However, our attempt to conceptualize VR as a “mediated third space” represents an analytical move: by systematically documenting language contact patterns in these emergent environments, we can inform intentional educational design for future VRSA programs. EN-JP exists within the digital wilds as an organic community. Whether the unintentional affordances that emerged in the observations (bilingual composition, ephemeral encounters, and role flexibility) can be deliberately replicated or enhanced in pedagogically designed VR worlds requires further investigation.

Unlike structured Collaborative Online International Learning (COIL) programs, which involve prearranged university partnerships, VR’s open-world platforms enable spontaneous encounters with globally distributed interlocutors. Language logs reflected this qualitative difference between teacher-mediated and peer-mediated practices (Theme 3, Table 2), with participants noting that VR provided their first extended conversations with nonteachers. VR’s embodied interaction further differentiates it from video conferencing, enabling gesture-based communication and spatial movement that approximate face-to-face exchange while remaining technologically mediated [16].

While our framework posits bilocation (see Theoretical Framework) as pedagogically productive, logs provide no evidence that students consciously experienced this dual positioning. None described drawing on their physical environment to enhance virtual interactions or reflecting on occupying two spaces simultaneously. Bilocation may

operate below conscious awareness. The concept is analytically useful for researchers, but in our data, it was phenomenologically unremarkable to learners experiencing VR as simply ‘another place to practice English. Similarly, while we theorized that iconic cues scaffold meaning-making, we did not code for iconic representation effects. Communication challenges occurred across modalities (Theme 1, Table 2), but it is unclear whether VR’s visual-spatial iconicity helped or hindered comprehension. While bilocation and iconicity offer compelling theoretical lenses, direct qualitative evidence of these phenomena in our data was limited; we position them as interpretive frameworks rather than empirically confirmed constructs in this study.

Kern et al.’s [17] framework on redistributed interaction factors proved more empirically tractable. Our findings document how VR’s technical interfaces (avatar-mediated identity, voice changers), social settings (ephemeral vs. persistent encounters), and spatial-temporal factors (deterritorialized access with technology barriers) redistributed contact opportunities toward self-initiated, peer-mediated, role-flexible interactions.

Collectively, these findings suggest that VR does not replicate the physical conditions of study abroad but instead creates qualitatively distinct learning environments. The marketplace framework explains *why* role reversal occurred (bilingual community composition shifts linguistic capital). The third-space framework explains *where* students were positioned (physically in Japan, virtually in EN-JP, mobilizing both L1 and L2). Kern’s interaction factors explain *how* platform affordances shape contact (avatar mediation, ephemeral encounters, autonomous seeking). However, our findings also reveal theoretical limitations: we cannot determine whether students consciously experienced bilocation as pedagogically productive, whether iconic representations scaffolded or hindered comprehension, or whether role reversals involved intercultural depth beyond functional code switching. Future research requires methods that capture not only frequency but also quality—such as conversation analysis, multimodal annotation, and stimulated recall—to test whether VR’s distinct ecology produces intercultural engagement or merely redistributes surface-level contact.

Across datasets, the pattern is consistent: VR shifts contact from instructor-led, institutionally structured, hierarchically stable exchanges toward self-initiated, peer-mediated, role-flexible exchanges. This shift explains both the spike in Japanese-as-resource episodes and the dip in overhearing and instructor contact. In short, VR does not thin the ecology as much as repatterning it.

Given that VR has the potential to redistribute contact opportunities while remaining theoretically underexamined, what practical guidance can we offer stakeholders seeking to implement or study VRSA programs?

6.5 Implications for stakeholders

The findings provide actionable guidance for educators, researchers, institutions, and platform designers working on supplemental VRSA programs.

For Educators VR can be used as a supplement or preparation for physical study abroad. Specify short, goal-focused sessions (e.g., 3 × 30 min/week) mixing roleplay service events, peer talk, and ambient listening; set a minimum time-on-task; pair activities with a micro-LCP diary for uptake tracking.

For Researchers The exposure windows are standardized across modalities, and the LCPs are reported together with session logs to distinguish between opportunity and

uptake. The LCP wording was adapted to distinguish scripted simulation/roleplay from live service encounters. Consider a synchronous/administered LCP variant for finer temporal resolution.

For Institutions Adequate device availability and low-friction usability should be ensured to reduce barriers to participation. Provide small hardware kits with a simple booking system, track usage to align with demand, and align assessments with exposure logs (micro-LCP) so that participation is visible, valid, and credit-bearing.

The implementation of VRSA programs requires both technological infrastructure and pedagogical alignment. In Japan, the broader concept of virtual ryūgaku encompasses various forms of online international exchange, from synchronous video conferences to asynchronous forum discussions. VRSA, as demonstrated in this study, represents the immersive VR subset of these approaches.

The practical implementation steps include the following:

- (1) Budget allocation: Initial investment of approximately ¥500,000 for 5–10 immersive VR headsets (with 6DoF tracking), with a ¥50,000 annual maintenance budget.
- (2) Staffing: Designing a VR coordinator for equipment management, orientation sessions, and technical support.
- (3) Scheduling: Offer VRSA as a 1–4 week supplementary program during semester breaks or as preparation for physical exchanges.
- (4) Credit framework: Award 1–2 credits on the basis of documented hours in VR (minimum 20 h) plus completion of language logs.
- (5) Safety protocols: Mandatory orientation covering motion sickness prevention, 30-minute session limits, and privacy guidelines are in place.
- (6) Partnership development: Seek populated online VR communities rather than creating isolated experiences. Connect students with existing VR language exchange communities (e.g., VRChat's EN-JP world).

Institutions can start with pilot programs involving 5–10 students to test feasibility before scaling up. VRSA works best as a complement to, rather than a replacement for, traditional exchanges, particularly for students who are unable to participate in physical study abroad programs due to financial or personal constraints.

For Platform Designers. To make the theoretical guidance concrete, we propose four enhancements, each mapped to one or more of the seven interaction factors and framed through a literacy-based design (genre, audience, mode, context, culture) [17]. In this way, the recommendations target specific semiotic resources so that platform choices systematically shape exposure, promote repeat encounters, and support relationship formation, optimizing second language learning.

- (1) Persistent ambient-audio layers: (language + spatial/temporal setting). Embed low-volume, location-specific 3-D soundscapes—convenience-store chimes, metro announcements, local music—into social VR worlds. Iconic audio cues signal place and time, whereas (privacy-safe) streamed snippets of authentic English conversation give learners more opportunities to overhear incidental dialog.
- (2) Scripted Nonplayer Character (NPC) service scenarios: (genre + social setting + culture). Offer on-demand nonplayer interactions that replicate everyday transactions (taxi ride, coffee-shop order, cafeteria food selection). Branching dialog and real-time

feedback enable users to rehearse goal-oriented, culturally grounded speech events when live opportunities are scarce.

- (3) Real-time gesture-and-pronunciation coaching: (technical interface + intersubjectivity). Provide an optional, low-salience overlay that surfaces brief cues when an avatar's hand movement (gestures), prosody, or proxemics approach a task-appropriate target exemplar. Keep feedback local and dismissible to avoid prescriptivism. Making tacit norms visible supports metacognitive reflection without disrupting flow and leverages VR-specific affordances rather than merely emulating physical study abroad.
- (4) Build Virtual Worlds for Repeat Encounters and Relationship Formation: (spatial-temporal settings + social settings: locality, routine, cohort stability). Physical study abroad often yields recurrent contact through mobility constraints (e.g., campus-bounded routines); in VR, unconstrained world hopping can disperse exposure. Intentional anchors—persistent home spaces, scheduled cohort meetups, and light friction on teleportation during sessions—are introduced to concentrate contact, deepen ties, and stabilize exposure patterns. Calibrate tasks and worlds to affordance profiles to elicit negotiation or ambient listening when needed, leveraging structure, interaction, and model [20].

While hardware and design constraints limit the speed at which platforms can adopt these features, implementing them would transform social VR from a generic “meeting room” into a richer, theory-aligned third-space environment, opening new avenues for VRSA.

7 Limitations and future directions

While our findings suggest the potential of VRSA, several methodological constraints limit generalizability and warrant a transparent discussion of these limitations. We note the main limitations—participation and technology, program structure, and measurement—as concrete threats to internal validity that future designs should directly address.

Technical and participation constraints. Despite VR's potential to democratize and enhance accessibility to study abroad, several practical constraints warrant consideration. Motion sickness remained an issue for some participants, even with modern VR features such as 6DoF headsets, which are designed to minimize such effects. Even though students who experienced adverse effects in VR during the orientation session were advised not to participate, one student reported experiencing intermittent nausea and limited VR use to periods when feeling well, showing how prolonged exposure can lead to discomfort even when brief orientations suggest tolerance. One participant who passed the initial screening later reported the following:

I experienced intermittent nausea after 20 min, limiting my practice to days when I felt completely well.

The fact that nausea occurred postorientation suggests cumulative sensitization effects. VR-induced nausea creates systematic bias and variable tolerance, potentially affecting time-on-task, as well as potentially introducing confounding technical literacy.

Individual differences also affect participation levels. Students reported varying amounts of time they could dedicate to VR because of competing academic

commitments and physical comfort limitations. One student noted an intention to be more active in VR, which was hampered by the demands of their schoolwork. These constraints underscore the need for flexible program design that accommodates individual differences while maintaining educational effectiveness.

Program structure and language exposure disparities. Language contact profile (LCP) surveys and language logs indicate significant structural differences between modalities that influence exposure patterns. In-person programs were structured and time-bound (approximately 3–6 weeks, mandatory), whereas VR was flexible and self-directed (approximately 1–4 weeks, voluntary). Students in the in-person programs reported English exposure more days per week than did those in VR, which is plausibly a function of program design rather than inherent modality effects. This limits direct comparability and may influence both the frequency and character of contact.

Measurement and sampling. The extreme effect sizes ($r = \pm 1.00$) observed between groups reflect genuine program differences rather than statistical artifacts. The complete separation of groups on ordinal scales, while unusual, is a legitimate outcome with a small sample when programs differ categorically.

The LCP involves several limitations. First, it relies on retrospective self-reports, introducing potential recall bias despite the use of weekly logs. Second, the instrument conflates simulated roleplay with authentic service encounters, a crucial distinction for VR communication platforms where scripted NPCs and live interactions coexist. Third, hour-based exposure estimates showed no significant between-group differences and were therefore deemphasized; however, this may obscure intensity differences within the reported contact days.

The unmodified LCP likely underestimates VR-specific exposure, as it conflates scripted role plays with live service encounters and omits avatar/embodiment affordances. We retained the original instrument for comparability but interpreted the findings as conservative.

The language logs, collected weekly in Japanese, yielded brief, practical entries rather than extended reflections, limiting our ability to provide rich narrative excerpts. However, the consistency of themes across participants suggests reliable patterns despite brevity. Modification of the LCP in future research could address these issues, allowing for within-subject sensitivity comparisons.

Sample size and practical constraints. Standard power approximations for comparing three groups ($\alpha = 0.05$, $1 - \beta \approx 0.80$) suggest that detecting medium-sized differences would require approximately 50–60 participants per group (150–180 total), whereas detecting the significant differences observed would require approximately 20–25 participants per group (60–75 total), depending on the variance structure and test family. With $N = 14$ across three groups, our study is underpowered for definitive inference—a limitation that we fully acknowledge.

Moreover, our design reflects the practical realities of study abroad research. The three programs examined here required institutional agreements spanning three countries, coordination across academic calendars, and substantial per-student costs. Partner institutions also have limited capacity, often accepting only 3–5 students per semester (with occasional larger short-stay cohorts), which constrains feasible participant sizes at any single site and time point.

For VR, the constraints are different but real. While headsets are reusable across cohorts, immersive VR (as used here) remains more costly than nonimmersive platforms (e.g., desktop or phone-based) and, in some circumstances, difficult to procure owing to production bottlenecks and release schedules.

Recent studies in both studies abroad and VR-for-language learning research underscore that primary studies commonly involve tens, not hundreds, of participants. Studies abroad include many single-site studies with a few dozen samples (large consortia inflate means; e.g., the Georgetown Consortium's $N \approx 1,300$) [36], and VR language meta-analyses likewise report typical per-study N s in the 10–100 range [37]. Participant numbers are also skewed, where immersive VR is positioned beside nonimmersive VR. These field norms help explain why (especially primary) studies rely on smaller sample sizes.

Accordingly, while our N is at the lean end, rather than abandoning such research entirely, we view this as exploratory evidence to guide program decisions and encourage larger investment in VRSA studies. The large effect sizes observed ($\epsilon^2 > 0.40$) are reported with robust summaries (means/medians/IQRs). While interpreted cautiously, given our sample size, the patterns suggest that meaningful differences exist between modalities that merit further investigation.

Future research should prioritize multi-institutional collaborations, pooling resources to achieve adequate statistical power while distributing costs across partners. Until such collaborations emerge, small-scale exploratory studies remain valuable for developing methodologies and generating hypotheses about virtual alternatives to traditional study abroad. We hope to continue collecting and analyzing longitudinal data year over year from sending students abroad. These limitations serve as a roadmap for scaling VRSA research globally; addressing them through multi-institutional collaborations will strengthen the evidence base for virtual reality study abroad as a viable complement to physical immersion programs.

Technology access and demographic constraints Our characterization of VR as ‘borderless’ and ‘deterritorialized’ requires qualification. VR removes geographic barriers and is often more affordable than physical study abroad (in the EFL context of Japan). Nevertheless, it remains costly for independent students without institutional support. Six-degree-of-freedom headsets cost between ¥40,000 and ¥80,000, requiring a stable income, robust technical infrastructure, and a certain level of digital literacy—these barriers shape who appears in VR language exchange communities. Our participants—engineering students at a technical college with institutional access to headsets—may have encountered interlocutors from similar socioeconomic and educational backgrounds, despite their geographic diversity.

Physical study abroad programs, while geographically bounded, potentially offer greater socioeconomic diversity through need-based scholarships, government exchange programs, and community-based homestays. VR’s ‘democratic access’ claim requires scrutiny: we may have traded one form of exclusivity (geographic/visa access) for another (technology/class access). Future research should examine the demographic composition of VR language communities to test whether they reproduce or disrupt existing privilege structures in study abroad access.”

Strategic implications and future research directions. Rather than negating VR's value, these constraints help define its strategic applications: providing accessible language contact when traditional immersion is cost-prohibitive or institutionally unfeasible and creating supportive environments for learners who benefit from lower-stakes practice opportunities.

Future research should prioritize several key areas:

- **Standardization:** Developing equivalent program structures across modalities, including standardized session lengths and learning objectives, to improve cross-study comparability.
- **Sample Diversity:** Recruiting significantly larger and more diverse cohorts to increase generalizability.
- **Program Design:** LCP data are used to design virtual study abroad programs that offer language contact opportunities on par with physical programs.
- **Interaction Opportunities:** Investigating the tendency for VR users to actively seek interaction with service personnel, potentially reflecting efforts to fulfill traditional study abroad expectations.

The observed role reversal, where Japanese students became 'native speaker resources' for others—something seen less in physical exchanges—warrants extensive future exploration to understand communication in hybrid environments.

8 Conclusion

In this exploratory pilot study, ($N = 14$), VRSA participants more often acted as linguistic resource providers, reporting less instructor contact and fewer incidental English encounters than did learners in physical programs. The Singapore participants presented the highest level of peer interaction. These patterns suggest that VR can complement and scaffold entry into physical, in-person study abroad exchanges.

Although the VRSA participants spent most of their weekly time in free conversation, they also sought more formal learning opportunities. Therefore, VR communication platforms could be improved for language learning by integrating different types of language exposure (e.g., structured instructor touchpoints alongside opportunities for ambient listening and simulated service encounters).

The limited time spent in VR and technical constraints (including software/hardware issues, as well as motion sickness) led to adverse effects for some students. While students across formats had broadly similar conversational opportunities, program "dosage" varied by context and availability. Language logs yielded valuable anecdotal insights but were not collected within a progressive qualitative framework.

Future research should (1) focus on recruiting larger and more diverse cohorts to increase generalizability. (2) Controlling and normalizing time-on-task across conditions spent in each context would help us better understand the differences in English exposure between programs. While the LCP asked students how much time (hours) they spent on each of these tasks, the differences between study abroad programs were not significant; therefore, they were excluded from the focus of this study. Further

studies could benefit from (3) implementing a more thorough qualitative analysis and (4) extending/innovating the LCP to understand how social media or smartphone use affects study abroad programs. For example, students can easily check their smartphones at any time. However, this is not possible when they use a head-mounted display (HMD). In addition, while this study focused on comparing VRSA (using an HMD) with in-person exchanges, (5) the addition of a teleconferencing condition (such as through Teams or Zoom) could help further explore the differences between online contexts. (6) Additional comparisons, such as screen recordings of VRSA interactions or videos of in-person exchanges, would enrich the present data and further help us understand the differences in the benefits of each condition.

VR shows strong potential to expand access and complement physical programs. As platforms add ambient audio layers and safer embodiment—and as programs standardize exposure and instructor touchpoints—VRSA can be evaluated not as a substitute but as a distinct ecology whose strengths are different from, yet synergistic with, in-person study abroad.

Appendix

See Tables 3 and 4.

Table 3 Rank-biserial correlations for pairwise comparisons following significant Kruskal–Wallis tests

Interaction type	BEL vs. SIN (<i>r</i>)	BEL vs. VR (<i>r</i>)	SIN vs. VR (<i>r</i>)
Talking with the instructor	0.11	0.83	1.00
Talking with classmates	-0.50	0.33	1.00
Conversations with service personnel	-0.56	0.53	0.87
Speaking Japanese with nonnative speakers	-1.00	-0.80	-0.07
Overhearing other people's English conversations	-1.00	-0.20	0.93

The values represent rank-biserial correlations ranging from -1 to +1

Negative values indicate that the first group had lower scores than the second group

Values of ± 1.00 indicate complete separation between groups with no overlapping scores, which can occur legitimately with small samples and ordinal data when program structures differ substantially

Table 4 Dwass–Steel–Critchlow–Fligner test statistics and p values for pairwise comparisons

interaction type	BEL vs. SIN (<i>W</i> , <i>p</i>)	BEL vs. VR (<i>W</i> , <i>p</i>)	SIN vs. VR (<i>W</i> , <i>p</i>)
Talking with the instructor	-0.500, 0.934	-3.727, 0.023	-3.651, 0.027
Talking with classmates	1.72, 0.442	-1.91, 0.365	-3.62, 0.028
Conversations with service personnel	1.86, 0.388	-2.28, 0.239	-2.94, 0.094
Speaking Japanese with nonnative speakers	3.905, 0.016	3.589, 0.030	0.213, 0.988
Overhearing other people's English conversations	3.90, 0.016	1.55, 0.517	-3.17, 0.065

Acknowledgements

The authors gratefully acknowledge the financial support from JSPS that made this study possible.

Author contributions

R.A.O. wrote the main manuscript text and prepared the table and figures. R.A.O. added the discussion and theoretical framework after the review process. D.T. organized participant groups and wrote descriptions of participants. D.T. added the language logs section upon review. All authors collaborated on arranging VR sessions. All authors reviewed the manuscript.

Funding

This study was supported by a Kiban-C Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) (Grant Number: 21K00778). The funding body played no role in the design of the study; the collection, analysis, or interpretation of the data; or the writing of the manuscript.

Data availability

Examples of LCP surveys and datasets generated and analyzed during the current study are available in the Zenodo repository, accessible via [\[https://zenodo.org/records/14851117\]](https://zenodo.org/records/14851117) (<https://zenodo.org/records/14851117>). Owing to privacy concerns, certain data are not publicly available but can be obtained from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the ethics committee of the National Institute of Technology, Hakodate College (KOSEN), as documented in the results of the NIT bioethics committee available at <https://www.hakodate-ct.ac.jp/disclosure/bioethics/> 令和6年度生命倫理審査結果. All procedures performed in the study were in accordance with the ethical standards of the institutional and national research committees and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent

was obtained from all participants included in the study. The participants provided consent for the publication of anonymized data.

Competing interests

The authors declare no competing interests.

Received: 4 March 2025 / Accepted: 20 January 2026

Published online: 27 January 2026

References

1. Moeller BA. Outbound student mobility in Japan nose-dived. Why does the government claim it went up? *Jpn Forum*. 2025;37(2):228–53. <https://doi.org/10.1080/09555803.2025.2457448>.
2. Esguerra JR, Figueroa RB, Rodriguez F, Fukuda A, Anzai Y. Collaborative online international learning: a reflexive thematic analysis of narratives from the metaverse. *Int J Inf Technol Gov Educ Bus*. 2024;6(1):100–8. <https://doi.org/10.32664/ijitgeb.v6i1.140>.
3. Chumpavan S, Boonyarattanasoontorn P, Mclver S, Tampanich S. Using technology to enhance Language learning in the digital era. *Engl Scholars Beyond Bord*. 2024;10(2):186–99.
4. Freed BF, Dewey DP, Segalowitz N, Halter R. The Language contact profile. *Stud Second Lang Acquis*. 2004;26(2):349–56. <https://doi.org/10.1017/S027226310426209X>.
5. Davidson DE, Garas N. What makes study abroad transformative? Comparing linguistic and cultural contacts and learning outcomes in virtual vs in-person contexts. *L2 J*. 2023;15(2):71–91. <https://doi.org/10.5070/L215260148>.
6. Grenby MO. Towards a history of children and heritage: young people, heritage education and the eighteenth-century 'Grand tour'. *Childh Past*. 2024;17(1):4–21. <https://doi.org/10.1080/17585716.2024.2331989>.
7. Twombly SB, Salisbury MH, Tumanut SD, Klute P. Special issue: study abroad in a new global century—Renewing the promise, refining the purpose. *ASHE High Educ Rep*. 2012;38(4):1–152.
8. Sasaki K, Uchiyama Y, Nakagomi S. Study abroad and the transnational experience of Japanese women from 1860s–1920s: four stages of female study abroad, Sumi Miyakawa and Tano Jōdai. *Espacio, Tiempo y Educación*. 2020;7(2):19–38. Available at: <https://dialnet.unirioja.es/descarga/articulo/7522576.pdf>
9. Isabelli-García C, Bown J, Plews JL, Dewey DP. Language learning and study abroad. *Lang Teach*. 2018;51(4):439–84. <https://doi.org/10.1017/S026144481800023X>.
10. Tseng WT, Liu YT, Hsu YT, Chu HC. Revisiting the effectiveness of study abroad Language programs: a multilevel meta-analysis. *Lang Teach Res*. 2021;25(1). <https://doi.org/10.1177/1362168820988423>.
11. DeKeyser R. Monitoring processes in Spanish as a second Language during a study abroad program. *Foreign Lang Ann*. 2010;43(1):80–92. <https://doi.org/10.1111/j.1944-9720.2010.01061.x>.
12. Kubota R. The social imaginary of study abroad: complexities and contradictions. *Lang Learn J*. 2016;44(3):347–57. <https://doi.org/10.1080/09571736.2016.1198098>.
13. Llanes À, Arnó E, Mancho-Barés G. Erasmus students using english as a lingua franca: does study abroad in a non-English-speaking country improve L2 english? *Lang Learn J*. 2016;44(3):292–303. <https://doi.org/10.1080/09571736.2016.1198097>.
14. Halenko N, Economidou-Koetsidis M. Thrivers and survivors during study abroad: the individual cases of Japanese learners of english. *Languages*. 2024;9(5):180. <https://doi.org/10.3390/languages9050180>.

15. Goodwin C. The co-operative, transformative organization of human action and knowledge. *J Pragmat.* 2013;46(1):8–23. <https://doi.org/10.1016/j.pragma.2012.09.003>.
16. Skarbez R, Smith M, Whitton MC. Revisiting milgram and kishino's reality-virtuality continuum. *Front Virtual Real.* 2021;2:647997. <https://doi.org/10.3389/frvir.2021.647997>.
17. Kern R, Liddicoat AJ, Zarate G. Research perspectives on virtual intercultural exchange in Language education. In: O'Dowd M, Lewis T, editors. *Virtual exchange for intercultural Language learning and teaching*. Oxfordshire: Routledge; 2022. pp. 1–20.
18. Kljajevic V. Spatial cognition in virtual reality. In: Kljajevic V, editor. *Consensual illusion: the mind in virtual reality*. Cognitive Systems Monographs, vol 44. Berlin, Heidelberg: Springer; 2021. pp. 113–134. https://doi.org/10.1007/978-3-662-63742-5_5.
19. Benna MK, Fusi S. Place cells May simply be memory cells: memory compression leads to Spatial tuning and history dependence. *Proc Natl Acad Sci U S A.* 2021;118(51):e2018422118. <https://doi.org/10.1073/pnas.2018422118>.
20. Aronov D, Tank DW. Engagement of neural circuits underlying 2D Spatial navigation in a rodent virtual reality system. *Neuron.* 2014;84(2):442–56. <https://doi.org/10.1016/j.neuron.2014.08.042>.
21. Bourdieu P. The economics of linguistic exchanges. *Soc Sci Inf.* 1977;16(6):645–68. <https://doi.org/10.1177/053901847701600601>.
22. EN-JP Language Exchange. Launch Instance—VRChat. 2024. Available from: https://vrchat.com/home/launch?worldId=w_rld_153be667-a86e-4aaf-9eed-921bd568ee9b.
23. VRChat Inc. VRChat [software]. VRChat Inc. 2024. Available from: <https://www.vrchat.com>.
24. Barricelli BR, Gadia D, Rizzi A, Marini DLR. Semiotics of virtual reality as a communication process. *Behav Inf Technol.* 2016;35(11):879–96. <https://doi.org/10.1080/0144929X.2016.1212092>.
25. Murgiano M, Motamedi Y, Vigliocco G. Situating Language in the real world: the role of multimodal iconicity and indexicality. *J Cogn.* 2021;4(1):38. <https://doi.org/10.5334/JOC.113>.
26. Tusher HM, Mallam S, Nazir S. A systematic review of virtual reality features for skill training. *Technol Knowl Learn.* 2024;29(2):843–78. <https://doi.org/10.1007/s10758-023-09713-2>.
27. Bigscreen. Inc. Bigscreen beta [software]. Bigscreen, Inc.; 2024. Available from: <https://www.bigscreenvr.com>.
28. ATI Studios. Mondly [mobile application software]. ATI Studios. 2014. Available from: <https://www.mondly.com>.
29. The jamovi project. jamovi (Version 2.6) [software]. 2024. Available from: <https://www.jamovi.org>.
30. Seliger HW. Does practice make perfect? A study of interaction patterns and L2 competence. *Lang Learn.* 1977;27(2):263–. <https://doi.org/10.1111/j.1467-1770.1977.tb00122.x>.
31. Nishida S. Working in Japanese where you are right now: internship program at cal'foscari university as a case study. *Proc Next Gener Glob Workshop.* 2021;14:1–29. (No DOI).
32. Dewey DP. Measuring social interaction during study abroad: quantitative methods and challenges. *System.* 2017;71:49–59. <https://doi.org/10.1016/j.system.2017.09.026>.
33. Long MH. Input, interaction, and second-language acquisition. *Ann N Y Acad Sci.* 1981;379(1):259–78. <https://doi.org/10.1111/j.1749-6632.1981>.
34. Lantolf JP, Thorne SL. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press. ISBN: 978-0194422499. (No DOI assigned by the publisher.).
35. Sauro S, Zourou K. What are the digital wilds? *Language Learn & Technol.* 2019;23(1):1–7. <https://doi.org/10.64152/10125/44666>.
36. Tseng W-T, Liu Y-T, Hsu Y-T, Chu H-C. Revisiting the effectiveness of study abroad Language programs: A multilevel meta-analysis. *Lang Teach Res.* 2024;28(1):156–200. <https://doi.org/10.1177/1362168820988423>.
37. Chen B, Wang Y, Wang L. The effects of virtual reality-assisted Language learning: A meta-analysis. *Sustainability.* 2022;14(6):3147. <https://doi.org/10.3390/su14063147>.
38. R Core Team. (2024). R: A Language and environment for statistical computing. (Version 4.4) [Computer software]. Available from <https://cran.r-project.org> (R packages retrieved from CRAN snapshot 2024-08-07).
39. Fox J, Weisberg S. (2023). car: Companion to Applied Regression. [R package]. <https://doi.org/10.32614/CRAN.package.car>.
40. Kerby DS. The simple difference formula: an approach to teaching nonparametric correlation. *Compr Psychol.* 2014;3., Article 1. <https://doi.org/10.2466/11.IT.3.1>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.