



# Multilingual language learning in a multimodal metaverse: a multidimensional study of communicative, affective, and cognitive development

Manuel Garcia    a,b,c

<sup>a</sup>College of Education, University of the Philippines Diliman, Quezon City, Philippines; <sup>b</sup>Educational Innovation and Technology Hub, FEU Institute of Technology, Manila, Philippines; <sup>c</sup>Graduate School of Education, Korea University, Seoul, South Korea

## ABSTRACT

**Introduction:** As digital platforms increasingly mediate language learning, the challenge is no longer simply how to deliver content online but how to design environments that cultivate authentic multilingual practice. While multilingualism has long been linked to enhanced metalinguistic awareness and domain-general cognitive flexibility, the role of multimodal digital environments in fostering these outcomes remains underexplored.

**Purpose:** Grounded in sociocognitive and multimodal interactionist perspectives, this study examines how a cross-device metaverse platform can support multilingual development through spatially organized, task-based, and avatar-mediated interaction. Specifically, it investigates whether multilingual engagement in language-zoned virtual spaces improves learners' communicative performance, affective engagement, and cognitive control compared to conventional instruction.

**Methodology:** Using a quasi-experimental cluster-assigned pretest-posttest control group design, learners engaged in communicative scenarios across English, Filipino, and Mandarin within language-zoned virtual spaces that cued role-appropriate language use. Data were collected using performance-based role-play assessments (code-switching accuracy, communicative competence), oral fluency measures (WPM), motivation and anxiety questionnaires, and a Stroop interference task to assess cognitive flexibility.

**Findings:** Compared to peers in a control condition, learners in the metaverse environment demonstrated significantly greater gains in code-switching accuracy, spoken fluency, motivational engagement, and cognitive control. Specifically, experimental participants showed improved context-appropriate language selection and reduced cross-language interference when shifting between English, Filipino, and Mandarin during task-based role-play scenarios. They also produced more fluent spoken output and demonstrated stronger communicative competence ratings in completing real-world interaction tasks. In addition, learners reported higher motivational engagement and cognitive results, further revealing improvements in inhibitory control and

## ARTICLE HISTORY

Received 25 July 2025

Accepted 11 January 2026

## KEYWORDS

Multimodal learning;  
language education;  
multilingualism; metaverse;  
avatar-mediated  
communication; educational  
technology

attentional regulation. Collectively, these outcomes suggest that spatially cued multilingual interaction in the metaverse supports integrated gains in linguistic performance and executive functioning. **Originality/Value:** This study provides empirical evidence that multilingual development is shaped not only by linguistic input but by how digital learning ecologies choreograph spatial, social, and multimodal cues into context-responsive language use. By operationalizing multilingual interaction through spatial language zoning, avatar-mediated tasks, and AI-supported multilingual dialogue, the study positions the metaverse as a semiotically rich pedagogical ecology that can simultaneously foster code-switching competence, oral fluency, motivational engagement, and domain-general executive control. The findings advance multimodal multilingual education theory by demonstrating how context-sensitive interaction design can generate co-emergent communicative, affective, and cognitive benefits in multilingual learners.

## Introduction

Multilingualism is increasingly recognized as a critical capability in today's linguistically diverse societies. Beyond its instrumental value in enhancing employability and cross-border mobility, multilingual proficiency contributes to cognitive flexibility, intercultural competence, and the ability to participate meaningfully in transnational social and professional networks. At a societal level, multilingual populations have been linked to greater civic inclusion, economic adaptability, and the preservation of linguistic heritage. However, the development of such competence remains a complex pedagogical challenge. While some learners acquire multiple languages organically in multilingual households or education systems, the majority contend with instructional models that emphasize grammatical accuracy over communicative adaptability. Traditional models grounded in monolingual ideologies also tend to isolate languages by task or classroom context (García and Wei 2014). As a result, learners are rarely given opportunities to develop cross-linguistic awareness or to practice dynamic code-switching in socially meaningful contexts. This fragmentation constrains the development of the metalinguistic awareness and interactional agility required to shift between codes in response to communicative demands (Spechtenhauser and Jessner 2024). These theoretical and practical tensions point to a critical need for pedagogical models that move beyond linear, language-by-language instruction and instead foster integrated, situated, and multimodal language use.

Responding to the need for more integrated and context-sensitive approaches to multilingual instruction requires a rethinking of the learning environment itself. The emergence of the metaverse as an avatar-based learning environment offers new possibilities for addressing the pedagogical limitations of traditional language instruction. Within these virtual worlds, learners can inhabit digital identities, interact within spatially organized contexts, and engage in socially situated communication that mirrors the complexity of multilingual encounters (Cantone et al. 2023; Lee 2023). Drawing on ecological perspectives of language learning (Gopalakrishnan 2022), the metaverse can be understood as an affordance-rich environment where meaning emerges through the interplay

of linguistic, visual, spatial, and embodied resources. Unscripted interactions shaped by context, role, and communicative intent foster a type of language use that is adaptive rather than formulaic. Language zones, shifting interlocutor dynamics, and task-based scenarios further create conditions that plausibly elicit code-switching and demand pragmatic negotiation across modalities. The spatialized design of these environments allows sociolinguistic norms to be embedded directly into virtual spaces and prompts learners to adjust their language practices in response to environmental and social cues. Understanding these affordances holds particular significance for multilingual education, as they foreground the potential of metaverses to operationalize multimodal forms of communication that more authentically mirror the dynamics of real-world linguistic interaction while cultivating intercultural awareness, adaptive language use, and cognitive flexibility.

## **Literature review**

### ***Language learning in immersive virtual environments***

Digital environments have increasingly been used to support second language acquisition (SLA), with prior research showing positive effects on learner autonomy, engagement, and performance (Tang 2024). Wong and Notari (2018) emphasized the value of computer-mediated reality in facilitating contextualized expression and sensory-rich interaction that support enhanced language practice. Lan (2020) and Xie et al. (2022) substantiated this claim by showing that immersive learning environments significantly improved motivation and self-regulation among learners studying English and Chinese. In basic education contexts, Lee, Yang, and Wu (2023) found that students using a virtual reality (VR) platform for language learning exhibited positive gains in behavioral, affective, and cognitive engagement, with improved performance in vocabulary post-tests. Similarly, recent studies within the Technology-Enhanced Language Learning (TELL) paradigm have demonstrated that metaverse-based instruction can foster intercultural communicative competence by promoting cultural awareness and adaptive communication in immersive contexts (e.g. Valizadeh and Morady Moghaddam 2025). These empirical studies are supported by a broader synthesis from Pérez-Jorge et al. (2025), whose systematic review confirmed that digital learning technologies significantly enhance vocabulary retention, engagement, and learner confidence in English as a second language (ESL) contexts. However, the review also noted that metaverses remain the least explored among emerging technologies in language education. While the expanding body of research on digital tools signals a promising trajectory for language learning, there remains a critical need to investigate how metaverse platforms can support this domain, particularly in advancing multilingual development.

### ***Multimodal communication in multilingual education***

The lack of research on metaverse-based language learning represents a missed opportunity, as one of its most pedagogically significant affordances is multimodal communication (Çelik and Baturay 2024). Multimodality refers to the ability to convey meaning through an integrated use of multiple semiotic resources that form the basis of human

interaction. Hasumi and Chiu (2024) emphasized that multimodality is central to emerging directions in language education, particularly in immersive environments where learners rely on more than just linguistic input to construct meaning. Multimodality is notably crucial in multilingualism as it enables learners to flexibly navigate between languages while using other modes to scaffold communication, negotiate meaning, and participate fully regardless of linguistic proficiency (Saint-Georges and Weber 2013). Lai (2024) provided empirical evidence that multimodal task design enhances both communicative competence and cognitive engagement in content and language integrated learning (CLIL) classrooms. Similarly, Jiang, Li, and Leung (2024) argued that digital multimodal composing as a form of translanguaging assessment in CLIL contexts allows multilingual students to integrate language with visual, auditory, and spatial elements in ways that more authentically represent their knowledge and identity. Nevertheless, even studies like Lee (2023), which employed a narrative-driven metaverse platform, fall short of adopting a multidimensional lens that considers critical facets of multilingual learning. There remains a need to examine how metaverses can support the development of multilingual competencies across multiple languages.

### ***A multidimensional view of multilingual and multimodal learning***

As multilingual learners begin to engage with increasingly multimodal digital spaces, reducing language learning to isolated outcomes risks overlooking the layered nature of communicative development. Translanguaging theory frames language learning as an inherently multidimensional process that draws simultaneously on linguistic, cognitive, affective, and interactional resources to enable flexible meaning-making across contexts (Wei 2018). Within this framework, translanguaging operates primarily as a sociolinguistic mechanism through which learners strategically mobilize their full linguistic repertoires in response to social roles, interactional demands, and communicative intent. Duarte (2020) demonstrated that translanguaging enables learners to integrate multiple languages and semiotic modes to support embodied and contextualized learning, which Tai and Wei (2024) further situated within CLIL settings as inherently multisensory and affective. Multimodality, by contrast, foregrounds the semiotic processes through which meaning is constructed across linguistic, visual, spatial, gestural, and embodied modes. From this perspective, multilingual performance involves the orchestration of diverse semiotic and linguistic practices, where learners fluidly navigate across named languages, communicative modes, spatial configurations, and affective stances to co-construct knowledge.

An ecological perspective integrates these dimensions by conceptualizing learning as emerging from the dynamic interaction between sociolinguistic practices (translanguaging), semiotic resources (multimodality), and the material–social environment in which communication unfolds. In the metaverse, these layers of communication converge as learners continuously navigate multimodal cues, linguistic choices, and social positioning to co-construct meaning (Cantone et al. 2023; Çelik and Baturay 2024). Such complexity underscores that multilingual meaning-making cannot be fully understood through a single linguistic, cognitive, or modal lens. Learners must negotiate semiotic resources to make sense of tasks. Overlooking any of these aspects jeopardizes rendering an incomplete picture of how learners meaningfully engage with multilingual content. Thus, a multidimensional lens

that foregrounds how learners interact, adapt, and make sense of multilingual tasks in digitally rich, socially complex environments is not only warranted but essential. These perspectives converge within an ecological view of multilingual learning, where translanguaging functions as the social-linguistic mechanism, multimodality as the semiotic process, and the metaverse as the contextual ecology that unites them. This synthesis clarifies how these frameworks collectively inform the study's multidimensional design.

### ***Gaps and research questions***

Despite the growing body of literature on TELL (Pérez-Jorge et al. 2025), significant gaps remain in understanding how metaverses can support multilingual learning in a multimodal environment. While prior research has demonstrated the positive impact of immersive learning spaces on engagement, vocabulary acquisition, and learner autonomy, most studies have focused on monolingual or bilingual contexts, and few have adopted a multidimensional lens that captures the interrelated communicative, affective, and cognitive processes involved in multilingual development. Although multimodality has been recognized as a key affordance of immersive environments (Hasumi and Chiu 2024), its role in fostering multilingual competencies has not been adequately explored. Moreover, the pedagogical potential of the metaverse in enabling embodied and socially mediated interaction across languages remains largely under-researched.

To address these converging gaps, the present study conceptualizes multilingual learning in the metaverse as a multidimensional phenomenon that engages learners' communicative behaviors, affective experiences, and cognitive processes. This framework acknowledges that linguistic performance (e.g. code-switching and fluency), emotional states (e.g. motivation and anxiety), and cognitive control (e.g. flexibility in task switching) are mutually supportive aspects of multilingual competence. From an ecological perspective, these dimensions are activated simultaneously by communicative environments that require learners to manage linguistic selection, attentional control, and affective regulation in real time. Accordingly, the study seeks to examine how participation in a multimodal metaverse environment may shape these interrelated dimensions of development without presupposing a unidirectional causal hierarchy among them. While each of these dimensions has been studied independently in prior research, the present study aims to provide an initial integrative exploration of how they converge within a metaverse-based multilingual environment rather than an exhaustive treatment of each. Guided by this objective, the study poses the following research questions:

- (1) Does participation in a multimodal metaverse environment improve students' code-switching proficiency compared to traditional bilingual instruction?
- (2) To what extent does the multimodal metaverse environment enhance students' speaking fluency and communicative competence in English, Filipino, and Mandarin?
- (3) What is the effect of metaverse-based multilingual interaction on students' language learning motivation and anxiety levels?
- (4) Does the metaverse-based multilingual environment have an impact on learners' cognitive flexibility as measured by a Stroop task?

## Methodology

### *Research design*

This study employed an experimental research approach to examine the effects of a multimodal metaverse-based learning environment on multilingual language acquisition across communicative, affective, and cognitive dimensions. Experimental research is a systematic method for investigating causal relationships by manipulating independent variables and observing their effects on dependent outcomes under controlled conditions. While true experiments rely on random assignment to ensure equivalence across groups, such control is often impractical in real-world educational contexts (Connolly, Ciara, and Urbanska 2018). Consequently, this study utilized a quasi-experimental design, which retains key features of experimental logic while accommodating naturally occurring classroom assignments. Specifically, a cluster-assigned pretest–posttest control group design was employed to facilitate both between-group comparisons (metaverse-based versus traditional instruction) and within-group comparisons across time (pre- to post-intervention). This design is particularly well-suited to applied educational research where internal validity must be balanced with ecological relevance. Ethical procedures were rigorously observed in accordance with both international standards and institutional review protocols.

### *Setting and participants*

This research forms part of a broader institutional initiative by a technological university to establish a fully digital campus built on a metaverse-based infrastructure. Situated in the capital city of the Philippines, the university is widely recognized as one of the pioneering institutions in the integration of metaverse technologies in higher education. English serves as the primary medium of instruction, while Filipino is the common mother tongue among students. Mandarin is offered as a foreign language course that introduces senior students to the foundational skills necessary for oral communication. The course emphasizes Chinese phonetic knowledge, pronunciation using pinyin, and provides exposure to cultural elements relevant to everyday conversational contexts. Two class sections enrolled in this course during the Academic Year 2024–2025 were recruited as study cohorts. A total of 80 students participated in the study, with 40 assigned to each group. Eligible participants were first-time enrollees in the Mandarin course with no prior formal instruction or native proficiency in the language and were undergraduate computing students aged 18–22 years. Given the university's English-medium instruction, all students possessed at least intermediate English proficiency, as verified through institutional placement records. While most students had previously encountered the university's metaverse environment through orientation, not all had extensive experience navigating it. Randomization was performed at the class-section level using a random number generator.

### *Multilingual and multimodal metaverse*

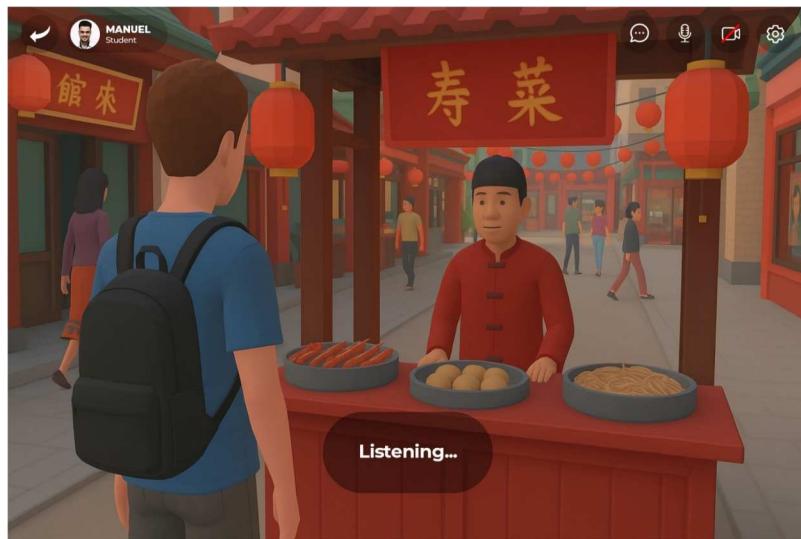
The experimental group engaged with a non-immersive metaverse platform designed to operate on widely accessible devices (e.g. smartphones, tablets, and laptops) without



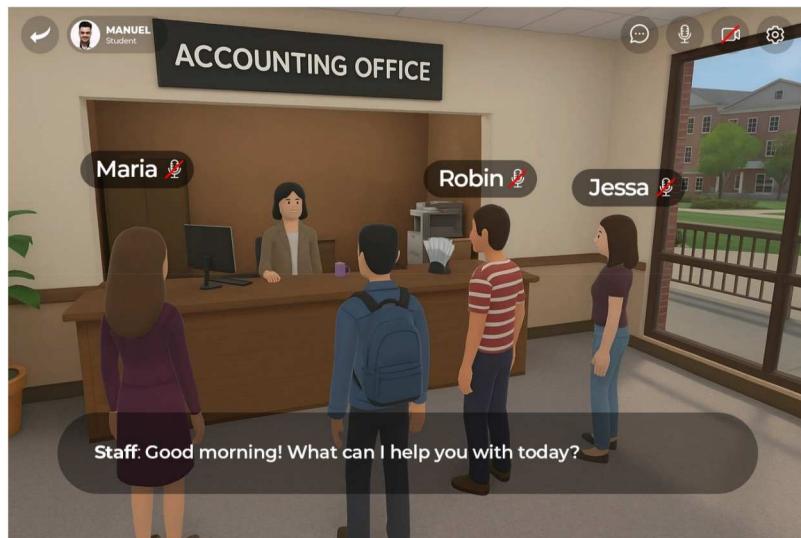
**Figure 1.** Stylized digital twin of a university campus within the metaverse platform.

requiring specialized virtual reality equipment. This configuration was intentionally selected to ensure accessibility, reduce technical barriers, and maintain ecological validity within the university's existing digital infrastructure, where students routinely access classes through standard devices. As emphasized in a recent systematic review, technical configuration often poses a significant barrier to implementing fully immersive VR environments in language classrooms (Parmaxi 2023). For the purposes of this study, the metaverse was customized to function as a stylized digital twin of a multicultural university environment (Figure 1). It mirrored real-world spatial and linguistic contexts to support scenario-based learning through task-oriented interactions within clearly defined zones that simulated authentic communicative challenges. Specifically, the virtual environment consisted of three primary zones, each associated with a designated instructional language. The Chinatown district (Figure 2) simulated a commercial area where interactions with non-playable characters (NPCs) occurred primarily in Mandarin. This zone required learners to complete tasks such as purchasing food, asking for prices, or responding to vendor inquiries using functional Mandarin expressions. In contrast, the campus area (Figure 3) designated English as the expected medium of communication, particularly for formal tasks such as submitting requests, seeking assistance, or clarifying academic concerns. Finally, the residential zone (Figure 4), modeled after student dormitories and communal living spaces, emphasized informal conversations in Filipino. Additional technical and interactional specifications are provided in Appendix A to give interdisciplinary readers a clearer understanding of the operational design.

Although each zone reinforced a specific language, contextual demands often prompted code-switching. For instance, a conversation that began in Filipino in the residential area could shift to English when discussing academic topics, or a dialogue in Mandarin within the Chinatown district might briefly transition to English for clarification. This intentional linguistic fluidity encouraged participants to adapt their speech according to social context, communicative intent, and interlocutor expectations. Participants also



**Figure 2.** Interaction with a vendor in Mandarin in the Chinatown District.



**Figure 3.** Requesting assistance in English at the Campus Accounting Office.

interacted with NPCs capable of engaging in context-sensitive conversations across all three languages. These characters were powered by a generative artificial intelligence model with multilingual capabilities to process learner inputs, infer communicative intent, and produce coherent responses in a language appropriate to the zone or evolving discourse cues. As the platform was designed as a multimodal environment, participants engaged through a combination of spoken language, text input, visual prompts, and interactive objects. Their avatars were also capable of performing expressive movements to complement and reinforce their verbal communication, either automatically triggered



**Figure 4.** Conversation with a classmate in Filipino in the Residential Area.

by system events or manually activated by users. For example, a participant might initiate a conversation by waving, nod to indicate understanding, or employ a questioning gesture to express uncertainty. The integration of verbal, visual, and gestural modes created a semiotically rich learning space aligned with multimedia learning and multimodal theories (Kress 2009; Mayer 2024).

### ***Experimental conditions and procedures***

Participants in the experimental group completed eight structured learning sessions over four consecutive weeks, each lasting 60 min and facilitated by the same instructor. Each session was built around a target scenario with increasing communicative complexity. Early sessions focused on foundational linguistic routines such as greetings, requests, and basic inquiries, while later sessions introduced problem-solving tasks requiring adaptive, multilingual responses (e.g. resolving miscommunications, handling dual-language requests, or mediating peer interactions). Brief pre-task instruction and vocabulary scaffolding were provided before each session, followed by a short reflective debriefing to consolidate learning.

To maintain alignment with the experimental group's linguistic exposure, the control group participated in equivalent 60-minute sessions delivered in a conventional classroom setting. Instruction included bilingual role-play exercises, targeted vocabulary drills, and guided sentence translation tasks designed around the same communicative themes as the metaverse-based sessions. For example, in weeks where the experimental group engaged in simulated service encounters, the control group practiced scripted customer-service dialogues and completed comprehension checks involving formal English responses and basic Mandarin expressions. Although the control condition lacked spatial or avatar-based context, all linguistic content was matched in terms of lexical scope, syntactic complexity, and discourse function. Both groups covered the same communicative

objectives, and pretest–posttest comparisons were conducted to evaluate differences in learning outcomes across the two instructional conditions.

All instructional sessions were conducted in alignment with the participants' established linguistic profile. As previously noted, participants were functionally bilingual in English and Filipino, in accordance with the definition of bilingualism as the ability to communicate effectively in both languages across academic and social domains. English served as the primary medium of instruction at the university, while Filipino functioned as the students' lingua franca in informal contexts. Mandarin was introduced as a foreign language for both groups. Both experimental conditions were delivered by the same instructor using parallel lesson plans, learning materials, and communicative targets to ensure instructional equivalence. The key distinction between the groups lay in the learning modality, with the experimental group interacting through the multimodal metaverse environment and the control group using alternative digital classroom tools such as video conferencing and presentation slides. A session-by-session breakdown of instructional activities, learning objectives, and language focus is outlined in Appendix B.

### ***Data collection and analysis***

Data collection involved pretest and posttest measures of code-switching proficiency, communicative competence, speaking fluency, language learning motivation and anxiety, and cognitive flexibility. Code-switching proficiency and communicative competence were both assessed through performance-based role-play tasks. For code-switching proficiency, participants completed timed scenarios requiring them to shift languages spontaneously in response to dynamic prompts. Each scenario was designed around functional communication tasks and developed based on the sociolinguistic frameworks of code-switching by Gumperz (1982), Poplack (1980), and Myers-Scotton (1993). Communicative competence was evaluated using the same video-recorded role-play performances, rated through a composite rubric measuring linguistic appropriateness, language choice, and pragmatic effectiveness, adapted from Canale and Swain (1980) and Canale (1983). Operational definitions and scoring descriptors for both constructs are presented in Appendix C. Speaking fluency was evaluated using structured oral narratives, scored for words per minute (WPM) and fluency markers, based on a rubric adapted from the Proficiency Guidelines by the American Council on the Teaching of Foreign Languages (ACTFL 2024). The rubric was modified to evaluate fluency in English, Filipino, and Mandarin, with attention to language-specific discourse norms. Affective constructs were measured using adapted versions of the Foreign Language Classroom Anxiety Scale (Horwitz, Horwitz, and Cope 1986) and the Language Learning Orientation Scale (Noels et al. 2000), both modified for multilingual contexts. Cognitive flexibility was measured using a metaverse version of the Stroop Color-Word Task (Figure 5), with the reaction time difference between incongruent and congruent trials serving as the interference score. Mixed-design ANOVAs were conducted to examine interaction effects between time (pretest, posttest) and group (experimental, control) on performance and Stroop outcomes. Paired-sample and independent-sample *t*-tests were used to compare within- and between-group differences for affective variables. Inter-rater reliability for speaking assessments was calculated using Cohen's kappa. Statistical significance was set at *p*



**Figure 5.** Students Taking Stroop Color-Word Tasks in the Metaverse.

<.05, and effect sizes were reported following second language research conventions (Plonsky and Oswald 2014).

## Results

All 80 participants completed the full four-week intervention and both pretest and posttest assessments, yielding a 100% retention rate and a complete dataset for analysis. No cases were excluded, and there were no instances of missing data across the measured variables. Preliminary screening confirmed that assumptions for normality and homogeneity of variance were met for all dependent measures, and no outliers exerted undue influence on the results. Descriptive and inferential statistics for all outcomes are summarized in Table 1, including group-wise means, standard deviations, within-group *p*-values, and associated effect sizes.

**Table 1.** Summary of pretest and posttest scores across outcomes.

Outcome Variable	Group	Pretest Mean (SD)	Posttest Mean (SD)	<i>p</i> -value (within-group)	Between-Group Effect
Code-Switching Accuracy (%)	Experimental	62.4 (8.3)	83.1 (7.2)	< .001	Significant ( $\eta^2$ = .18)
	Control	63.1 (7.9)	66.2 (8.1)	.112	
Speaking Fluency (WPM)	Experimental	85.2 (10.1)	102.6 (9.4)	.004	Significant ( $\eta^2$ = .14)
	Control	84.9 (9.7)	88.5 (10.5)	.091	
Communicative Competence (1–5)	Experimental	2.8 (0.6)	4.1 (0.5)	< .001	Significant ( $d$ = 1.01)
	Control	2.9 (0.7)	3.2 (0.6)	.087	
Motivation (1–7 Likert)	Experimental	4.9 (0.8)	5.8 (0.6)	.001	Significant ( $d$ = 0.85)
	Control	5.0 (0.7)	5.1 (0.8)	.298	
Anxiety (1–7 Likert)	Experimental	4.7 (0.7)	4.4 (0.8)	.068	Not Significant
	Control	4.6 (0.6)	4.5 (0.7)	.355	
Stroop Interference (ms)	Experimental	312 (45)	268 (37)	.021	Significant ( $\eta^2$ = .09)
	Control	309 (47)	304 (44)	.601	

Note: Effect sizes are reported as partial eta squared ( $\eta^2$ ) for outcomes analyzed using mixed-design ANOVAs and as Cohen's  $d$  for pairwise t-test comparisons. Partial  $\eta^2$  reflects the proportion of variance attributable to the interaction, while Cohen's  $d$  represents the standardized mean difference between pretest and posttest scores within each group.

***RQ1. does participation in a multimodal metaverse environment improve students' code-switching proficiency compared to traditional bilingual instruction?***

A mixed-design ANOVA revealed a significant interaction between time (pretest, posttest) and group (experimental, control) on code-switching proficiency,  $F(1, 78) = 12.47, p < .001$ , partial  $\eta^2 = .18$ , indicating that gains over time differed as a function of instructional condition. Follow-up pairwise comparisons using Bonferroni correction revealed that the experimental group demonstrated a statistically significant increase in code-switching accuracy from pretest ( $M = 62.4, SD = 8.3$ ) to posttest ( $M = 83.1, SD = 7.2$ ),  $t(39) = 8.12, p < .001$ . In contrast, the control group exhibited a non-significant change over the same period,  $t(39) = 1.04, p = .31$ . These results indicate that learners in the metaverse-based condition showed substantial improvement in managing language alternation. The integration of spatial and linguistic cues appears to have supported more context-sensitive code selection and adaptive bilingual switching in real time.

***RQ2: to what extent does the multimodal metaverse environment enhance students' speaking fluency and communicative competence in English, Filipino, and Mandarin?***

For speaking fluency, a significant main effect of time was detected,  $F(1, 78) = 9.02, p = .004$ , partial  $\eta^2 = .14$ . The mean WPM score of the experimental group increased from  $M = 85.2, SD = 10.1$  at pretest to  $M = 102.6, SD = 9.4$  at posttest, whereas the control group's change from  $M = 84.9, SD = 9.7$  to  $M = 88.5, SD = 10.5$  did not reach statistical significance. With respect to communicative competence, the experimental group showed a significant improvement,  $t(39) = 6.77, p < .001$ , with scores increasing from  $M = 2.8, SD = 0.6$  to  $M = 4.1, SD = 0.5$  and a corresponding large effect size ( $d = 1.01$ ). The control group's gains were not statistically significant,  $t(39) = 1.22, p = .23$ . These findings suggest that avatar-mediated interaction in the metaverse promoted greater fluency and pragmatic control than conventional classroom instruction. The spatially anchored communicative zones likely provided meaningful situational cues that scaffolded discourse planning and turn-taking across multiple languages.

***RQ3: what is the effect of metaverse-based multilingual interaction on students' language learning motivation and anxiety levels?***

Motivation scores in the experimental group increased significantly from pretest ( $M = 4.9, SD = 0.8$ ) to posttest ( $M = 5.8, SD = 0.6$ ),  $t(39) = 3.98, p = .001$ , yielding a large effect size ( $d = 0.85$ ). However, although anxiety scores declined modestly from  $M = 4.7, SD = 0.7$  to  $M = 4.4, SD = 0.8$ , the reduction was not statistically significant,  $t(39) = -1.89, p = .068$ . No significant changes in motivation or anxiety were observed in the control group (motivation:  $t(39) = 1.06, p = .298$ ; anxiety:  $t(39) = -0.94, p = .355$ ). These outcomes indicate that engagement within the metaverse environment heightened learners' motivational investment while concurrently maintaining manageable anxiety levels. The socially interactive yet low-pressure nature of avatar-mediated exchanges may have contributed to a more affectively supportive learning environment.



#### ***RQ4: does the metaverse-based multilingual environment have an impact on learners' cognitive flexibility as measured by a Stroop task?***

A statistically significant interaction between time and group was found for Stroop interference scores,  $F(1, 78) = 5.63, p = .021$ , partial  $\eta^2 = .09$ . Participants in the experimental condition exhibited a reduction in reaction time difference between incongruent and congruent trials, decreasing from  $M = 312$  ms,  $SD = 45$  to  $M = 268$  ms,  $SD = 37$ . The control group's corresponding difference scores remained stable across testing sessions ( $M = 309$  ms,  $SD = 47$  to  $M = 304$  ms,  $SD = 44$ ),  $t(39) = 0.52, p = .601$ . This pattern demonstrates a measurable improvement in inhibitory control among learners exposed to the metaverse condition. Sustained engagement with concurrent multimodal and multilingual cues appears to have strengthened cognitive mechanisms underlying attentional switching and interference management.

### **Discussion**

The growing imperative to cultivate multilingual proficiency in increasingly complex communicative landscapes has underscored the need for pedagogical environments that reflect the sociolinguistic realities of learners. While multimodal learning has long been recognized as a critical enabler of embodied language use, the affordances of the metaverse as a socially and spatially contextualized environment have yet to be fully leveraged in multilingual pedagogy. This study addresses a persistent gap in the literature by operationalizing translanguaging, interactional competence, and multimodal engagement within a virtual ecology. It responds to the absence of multidimensional, ecologically valid investigations into how metaverse-based environments support the communicative, affective, and cognitive facets of multilingual development. Empirical results indicate that a multilingual and multimodal metaverse environment can significantly enhance learners' code-switching proficiency, speaking fluency, communicative competence, and cognitive flexibility. Motivational gains further underscore the platform's affective resonance, although reductions in anxiety did not reach statistical significance. Collectively, these findings establish the viability of metaverse environments as theoretically grounded tools for language education and empirically substantiate the premise outlined in the literature review that metaverse-based learning can integrate multimodal, multilingual, and affective processes within a single pedagogical ecology.

#### ***Code-switching as situated competence in the multilingual metaverse***

Interestingly, the significant gains observed in code-switching accuracy among metaverse learners call into question long-held assumptions that frame code-switching as a compensatory linguistic strategy. While translanguaging foregrounds the fluid integration of linguistic repertoires for meaning-making, code-switching in this study is treated as an interactionally cued act within that broader translanguaging continuum. Within the metaverse, code-switching did not occur as fallback behavior but emerged as a situated form of communicative competence activated by social and spatial contingencies. This finding reinforces earlier sociolinguistic claims that code-switching, particularly intra-sentential shifts, demands a high level of grammatical and pragmatic control (Yim and Clément

2021). Rather than merely compensating for linguistic gaps, learners in the metaverse were actively navigating layered communicative expectations shaped by spatial zones, social roles, and task contingencies. Thus, the study meaningfully extends the literature on translanguaging and multilingual pedagogy by operationalizing code-switching as an interactionally necessary act within a designed digital ecology. While earlier studies have documented learners' code-switching in naturalistic or classroom-based settings, few have scaffolded this phenomenon through intentional environmental cues. To illustrate how these adaptive shifts unfolded in practice, Appendix D presents representative learner-NPC exchanges drawn from each metaverse zone showing how participants negotiated meaning through context-sensitive code-switching. Unlike prior interventions that framed code-switching as a post hoc linguistic outcome, this study treated it as a socio-pragmatic act deeply embedded in the learner's navigation of spatialized norms, role expectations, and communicative goals. In this regard, the results are consistent with the argument that code-switching can serve to index group belonging, project identity, and signal pragmatic nuance (Bahous, Baroud, and Bacha 2014; Gardner-Chloros 2009).

What is particularly noteworthy is the magnitude of improvement, which was not mirrored in the control group despite comparable lexical exposure. This deviation suggests that the learning gains were not simply a function of language input but were catalyzed by the dynamic interplay of environmental cues, task-based demands, and avatar-mediated interaction (Cantone et al. 2023). Participants transitioned fluidly between languages not because they were told to do so, but because the spatial architecture and task constraints made it pragmatically necessary. These findings support an ecological view of language learning, wherein linguistic behavior emerges in response to the semiotic, spatial, and social cues embedded in the environment (Gopalakrishnan 2022; Steffensen and Kramsch 2017). One likely reason why these results diverged from earlier studies lies in the specificity of environmental scaffolding. Previous platforms often treated the environment as a neutral backdrop by relying on pre-scripted dialogue or linear activities (e.g. Göbel et al. 2024). In contrast, this study employed AI-powered NPCs and a spatial language-zoning system that cued learners toward code-switching through contextual necessity. These features mirrored real-world multilingual encounters where shifts in code are often conditioned by role, topic, and space (Bahous, Baroud, and Bacha 2014). Overall, these insights reposition code-switching as a context-sensitive strategy that reflects multilingual expertise.

### ***Spatial design and the emergence of multilingual fluency***

Whereas code-switching was driven by moment-to-moment role negotiation and linguistic necessity, fluency emerged from the repeated alignment between spatial structure, task demands, and communicative expectation. The observed improvements in speaking fluency and communicative competence among learners in the metaverse condition reinforce the view that spatial configuration is a generative semiotic framework that actively organizes linguistic behavior. This perspective resonates with Christou, Parmaxi, and Christoforou (2025), whose systematic review on extended reality (XR) for language learning identified spatial immersion, multimodal literacy, and authentic task alignment as key contributors to oral proficiency gains. The present study builds on

this foundation by demonstrating that even in less immersive digital environments than XR, spatially zoned architecture can elicit contextually responsive speech through continuous interaction with visual, gestural, and procedural cues. These spatial arrangements operated as instructional scaffolds that shaped how and when language was used (Bacca-Acosta et al. 2022; Lazovic 2025). In this sense, fluency was not simply the result of increased exposure to language-rich environments, but the outcome of learners repeatedly responding to pragmatically motivated demands embedded in their virtual surroundings. This dynamic foreground spatial-task alignment as a central mechanism of fluency development, where language is not rehearsed in abstraction but enacted through situationally anchored participation (Marre, Nathalie, and Labeye 2024).

While prior research on VR-supported language learning often emphasized the benefits of immersive presence or motivational novelty, such accounts have typically under-theorized the pedagogical role of spatial structure in shaping discourse production. In contrast, the current findings suggest that spatial zoning functions as a discourse-organizing mechanism that prompts learners to modulate speech in accordance with environmental affordances. This interpretation refines earlier claims by Žammit (2023), who highlighted the potential of VR to support minority language acquisition through culturally situated scenarios, by specifying how spatially encoded expectations can drive fluency across multiple languages within a unified interface. Rather than simulating real-world locations for their aesthetic or affective value, the metaverse environment used in this study operationalized spatial logic as a pedagogical cue system, where movement across zones triggered shifts in register, role, and linguistic rhythm. These findings align with those of Çelik and Baturay (2024), who demonstrated that metaverse-based environments can significantly enhance vocabulary retention and classroom community by integrating task-based language learning within avatar-mediated settings. Such convergence suggests that spatial fluency may emerge not from immersion per se, but from the learner's ability to interpret and act upon semiotic cues that organize speech temporally, socially, and pragmatically. These insights support a view of fluency as an emergent capacity cultivated through goal-oriented interaction with semiotic landscapes that are pedagogically constructed to elicit purposeful language use.

### ***Motivation outpaces anxiety in avatar-based language learning***

Contrary to the common assumption that increased motivation in language learning necessarily entails a corresponding decrease in anxiety, the present findings reveal a more nuanced affective landscape. While learners in the metaverse condition demonstrated a pronounced surge in motivational engagement, their anxiety levels remained largely unchanged. This finding suggests that these affective states are mediated by distinct cognitive and environmental triggers. This bifurcation reflects emerging perspectives in affective-cognitive research that caution against viewing motivation and anxiety as binary opposites (MacIntyre and Gregersen 2012). In the current study, motivational gains appeared to stem from the interplay of avatar embodiment, spatial interactivity, and task-directed autonomy. Learners interacted with the environment through avatars that allowed them to offload self-consciousness onto a digital proxy. According to prior works (e.g. Hu et al. 2023), this mechanism has been shown to increase task focus and reduce identity threat. Moreover, the immersive architecture of the platform

scaffolded meaningful participation by embedding language tasks within socially and cognitively purposeful scenarios. This approach is supported by Ukenova et al. (2025), who demonstrated that emotionally expressive avatars can significantly enhance learners' affective investment and sense of agency, even in early-stage systems with limited instructional precision.

Yet despite the motivational uplift, anxiety levels in the experimental group did not exhibit a significant decrease. This finding complicates prior narratives suggesting that digital or immersive environments inherently reduce foreign language anxiety (Thrasher 2022; York et al. 2021). While avatar-mediated interaction may buffer learners from overt judgment, the synchronous, real-time nature of dialogue with NPCs in the current study likely sustained a form of performance pressure akin to face-to-face communication. This interpretation aligns with Ukenova et al.'s (2025) findings, where the novelty and emotional interactivity of avatars initially elevated learner interest, but did not consistently translate into reduced apprehension in contexts requiring spontaneous production and public output. In contrast to high-immersion VR studies such as Kaplan-Rakowski and Gruber (2023), which reported significant reductions in foreign language anxiety through repeated public speaking simulations using fully immersive headsets and embodied avatars, the non-immersive yet dialogically demanding design of the present metaverse experience may have preserved cognitive strain, especially for learners unaccustomed to open-ended performance in multilingual settings. Although anxiety was assessed globally rather than by specific task or language, observational notes suggested that apprehension tended to peak during spontaneous Mandarin exchanges and real-time clarification tasks with NPCs, where cognitive load and linguistic unfamiliarity converged. In contrast, interactions in English and Filipino zones appeared more routinized and less affectively charged. These inferred patterns underscore that anxiety within multilingual metaverse contexts may vary as a function of task complexity, linguistic distance, and immediacy of response demands. Therefore, the findings suggest that motivation and anxiety operate as co-existing yet independently modulated affective systems within avatar-based language environments, such that while motivation may be readily activated through design features that foster presence and self-direction, anxiety may demand longer-term exposure, affective scaffolding, and differentiated interaction pacing.

### ***Multimodal interaction enhances cognitive control across languages***

Few aspects of language learning reveal the cognitive depth of multilingualism as clearly as the management of interference in multimodal tasks. In this study, participants exposed to a metaverse-based multilingual environment demonstrated significantly reduced Stroop interference scores, suggesting that navigating linguistically and visually layered spaces may enhance executive control processes, particularly those related to inhibition and attentional flexibility. These results expand on earlier work by Marian et al. (2013), who found that multilinguals demonstrate greater within-language interference relative to between-language interference as a function of proficiency and language configuration. The present findings add a crucial layer by indicating that in environments where linguistic, spatial, and visual modes are tightly integrated, inhibitory control develops through both language exposure and the dynamic coordination of semiotic inputs. This interpretation aligns with Van Heuven et al. (2011), who showed that Stroop interference in trilinguals

varies based on orthographic similarity and language pairings, reinforcing the idea that cross-modal and cross-linguistic coordination intensifies cognitive demands. It also resonates with findings by Acha-Amankwaa et al. (2023), who observed that multilingual older adults exhibited domain-general gains in interference suppression, even when controlling for demographic and experiential variables. Together, these studies support the view that executive control is not merely a byproduct of bilingualism but a function of how language is operationalized across modalities and task ecologies.

Notably, this outcome challenges assertions that the so-called bilingual cognitive advantage lacks robustness in controlled samples (Kousaie and Phillips 2012; Ware, Kirkovski, and Lum 2020). While prior studies have questioned whether bilingualism alone confers measurable gains in executive function, the present study underscores that such effects may be more clearly observed in multilingual learners operating within cognitively demanding environments. Freeman et al. (2022) similarly emphasized that Stroop performance is shaped less by bilingual status per se than by the sociolinguistic and cognitive characteristics of the interactional context. In this study, participants engaged in continuous language selection and interference suppression across multiple linguistic repertoires, as the use of three languages was activated by spatial, social, and task-driven cues. These demands embedded in avatar-mediated, multimodal exchanges appear to foster executive control through ecological task engagement rather than through language proficiency alone. Intriguingly, the cognitive benefits observed here stem not from the presence of multiple languages in isolation, but from the multilingual learner's need to manage them simultaneously within an orchestrated communicative ecology. Consequently, the present study affirms that carefully designed multimodal environments can serve as effective training grounds for cognitive control, with implications that extend beyond language acquisition to broader domains of neurocognitive adaptability and instructional design.

### ***Theoretical, pedagogical, and technical design implications***

This study contributes to the evolving theoretical landscape of multilingual education by empirically demonstrating how code-switching, communicative competence, affective engagement, and cognitive flexibility can be understood as co-emergent phenomena within spatially and socially situated environments. Drawing on translanguaging theory, ecological linguistics, and multimodal learning frameworks, the findings reframe language acquisition as a fundamentally situated, performative, and adaptive process. Rather than viewing multilingual development as a linear accumulation of discrete language competencies, this study theorizes it as the orchestration of semiotic resources across spatial, social, and interactional dimensions. By embedding linguistic decision-making within the temporal flow of authentic tasks, the research challenges static models of multilingualism and calls for new theoretical paradigms that account for the embodied, context-responsive nature of language use in digital ecologies.

From a pedagogical perspective, the findings constitute a decisive critique of traditional and monolingually framed language instruction. The substantial improvements in fluency and code-switching accuracy observed in the metaverse condition suggest that pedagogical designs grounded in spatial differentiation, role-based immersion, and avatar-mediated interaction can elicit more authentic and cognitively engaging

language use than conventional classroom formats. Crucially, learners were not instructed when or how to switch codes. Instead, they responded to the affordances of the environment and demonstrated a level of pragmatic sensitivity often absent in scripted educational settings. This aspect repositions the metaverse learning environment as an active co-constructor of linguistic behavior. Furthermore, the dissociation between heightened motivation and persistent anxiety underscores the importance of treating affect not as a unidimensional construct but as a composite of co-existing states that respond to different environmental triggers. Effective multilingual pedagogy in digital environments must consequently integrate both motivational catalysts and affective scaffolds. This approach tailors the design to support learners cognitively and emotionally during high-stakes interaction.

Technologically, the study makes a compelling case for the promising potential of non-immersive and cross-device metaverse platforms as scalable and research-driven environments for multilingual instruction. Despite the absence of head-mounted displays or motion-tracking systems, the platform delivered measurable cognitive and linguistic gains by leveraging carefully constructed affordances such as language-zoned areas, intelligent NPCs, and gesturally responsive avatars. These features functioned not as cosmetic enhancements but as pedagogical mechanisms that translated theoretical principles into actionable interaction design. Embedding expectations for language use directly into spatial organization and task progression allowed the environment itself to serve as a dynamic engine for learning. This quality signals an important shift in TELL and educational technology in general. Rather than acting as external tools that support instruction, these metaverse environments embody pedagogical intent through their very structure. For designers and developers, this recognition highlights the importance of prioritizing contextual realism, semiotic richness, and adaptive learner interaction as foundational design principles that determine the educational efficacy of digital learning spaces.

### ***Limitations and future research***

While this study offers compelling empirical support for the didactic value of multimodal metaverses in multilingual language learning, several limitations warrant consideration.

First, although the quasi-experimental design is well suited to real-world educational contexts, it constrains internal validity. The absence of individual-level randomization limits control over potential confounding factors such as baseline digital fluency, intrinsic motivation, or prior multilingual exposure. Future research should employ more rigorously controlled designs to strengthen causal inference and clarify the mechanisms underlying observed outcomes.

Second, the intervention was implemented through a non-immersive metaverse accessed via standard laptops and mobile devices. While this configuration enhanced accessibility and yielded notable learning gains, it restricts conclusions regarding embodied cognition and spatial presence. Comparative studies examining varying levels of immersion are needed to determine how embodiment influences language switching, attentional control, and affective dimensions.

Third, the study focused primarily on spoken interaction, which overlooked other dimensions of multilingual literacy such as collaborative writing, reading comprehension,

and digital translanguaging. Future work should broaden assessment to include written and receptive skills and integrate qualitative approaches (e.g. interactional discourse analysis, multimodal gesture tracking, and avatar movement analysis) to better capture how learners coordinate linguistic and non-linguistic resources in complex communicative contexts.

Finally, while the study was situated within a trilingual context in which English, Filipino, and Mandarin occupy distinct educational and sociolinguistic functions, the transferability of these findings to other multilingual configurations warrants careful consideration. Linguistic ecologies characterized by different power relations or language ideologies (e.g. Arabic–French–English settings or minority–majority pairings) may produce distinct code-switching behaviors, identity alignments, and affective responses. Future research should investigate how metaverse-based multilingual instruction operates within such hierarchically or ideologically stratified settings to determine which pedagogical affordances remain stable and which require contextual adaptation.

These limitations do not detract from the study's contributions but instead point to promising future directions at the intersection of multilingualism, multimodality, and virtual learning. Continued research across these domains is essential for developing theoretically grounded frameworks for multilingual instruction in next generation learning environments.

## Conclusion

The metaverse presents a promising frontier for multilingual education by enabling multimodal learning to unfold within socially interactive and spatially organized digital ecologies. The present study demonstrates that embedding language learning in such multidimensional contexts fosters measurable gains in code-switching accuracy, communicative fluency, cognitive control, and motivational engagement. Learners who navigated the multilingual metaverse showed notable improvements in real-time language switching, produced more fluent and pragmatically appropriate speech, and exhibited enhanced inhibitory control on Stroop interference tasks while reporting greater motivation without increased anxiety. The findings reported here provide empirical evidence that metaverses can cultivate complex language competencies by simulating the communicative demands and semiotic diversity of real-world multilingual interaction. As digital learning ecosystems continue to evolve, this research advances current understanding of how multimodal and multilingual environments can serve as scalable and cognitively enriching spaces for developing multidimensional language competence.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

*Manuel Garcia*  <http://orcid.org/0000-0003-2615-422X>

## References

Acha-Amankwa, P., E. Kushnereva, H. Miksch, J. Stumme, S. Heim, and M. Ebersbach. 2023. "Multilingualism is Associated with Small Task-specific Advantages in Cognitive Performance of Older Adults." *Scientific Reports* 13 (1): 1–11. <https://doi.org/10.1038/s41598-023-43961-7>.

ACTFL. 2024. *ACTFL Proficiency Guidelines* 2024. [https://www.actfl.org/uploads/files/general/Resources-Publications/ACTFL\\_Proficiency\\_Guidelines\\_2024.pdf](https://www.actfl.org/uploads/files/general/Resources-Publications/ACTFL_Proficiency_Guidelines_2024.pdf).

Bacca-Acosta, J., J. Tejada, R. Fabregat, Kinshuk, and J. Guevara. 2022. "Scaffolding in Immersive Virtual Reality Environments for Learning English: An Eye Tracking Study." *Educational Technology Research and Development* 70 (1): 339–362. <https://doi.org/10.1007/s11423-021-10068-7>.

Bahous, R. N., N. M. Baroud, and N. N. Bacha. 2014. "Code-Switching in Higher Education in a Multilingual Environment: A Lebanese Exploratory Study." *Language Awareness* 23 (4): 353–368. <https://doi.org/10.1080/09658416.2013.828735>.

Canale, M. 1983. "From Communicative Competence to Communicative Language Pedagogy." In *Language and Communication*, edited by J. C. Richards and R. W. Schmidt, 2–27. Longman. <https://doi.org/10.4324/9781315836027-2>.

Canale, M., and M. Swain. 1980. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing." *Applied Linguistics* 1 (1): 1–47. <https://doi.org/10.1093/applin/1.1.1>.

Cantone, A. A., R. Francese, R. Sais, O. P. Santosuosso, A. Sepe, S. Spera, and G. Vitiello. 2023. "Contextualized Experiential Language Learning in the Metaverse." Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter. <https://doi.org/10.1145/3605390.3605395>.

Celik, F., and M. H. Baturay. 2024. "The Effect of Metaverse on L2 Vocabulary Learning, Retention, Student Engagement, Presence, and Community Feeling." *BMC Psychology* 12 (1): 1–17. <https://doi.org/10.1186/s40359-024-01549-4>.

Christou, E., A. Parmaxi, and M. Christoforou. 2025. "Implementation and Application of Extended Reality in Foreign Language Education for Specific Purposes: A Systematic Literature Review." *Universal Access in the Information Society* 24 (3): 2061–2076. <https://doi.org/10.1007/s10209-025-01191-w>.

Connolly, P., K. Ciara, and K. Urbanska. 2018. "The Trials of Evidence-Based Practice in Education: A Systematic Review of Randomised Controlled Trials in Education Research 1980–2016." *Educational Research* 60 (3): 276–291. <https://doi.org/10.1080/00131881.2018.1493353>.

Duarte, J. 2020. "Translanguaging in the Context of Mainstream Multilingual Education." *International Journal of Multilingualism* 17 (2): 232–247. <https://doi.org/10.1080/14790718.2018.1512607>.

Freeman, M. R., J. J. Robinson Anthony, V. Marian, and H. K. Blumenfeld. 2022. "Individual and Sociolinguistic Differences in Language Background Predict Stroop Performance." *Frontiers in Communication* 7:1–18. <https://doi.org/10.3389/fcomm.2022.865965>.

García, O., and L. Wei. 2014. *Translanguaging: Language, Bilingualism and Education*. London: Palgrave Pivot. <https://doi.org/10.1057/9781137385765>.

Gardner-Chloros, P. 2009. *Code-Switching*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511609787>.

Göbel, K., S. Lars, N. Katharina, and L. Struck. 2024. "Appreciation of Multilingual Teaching Activities by Secondary School Students in Germany: Findings from a Quasi-experimental Intervention Study on Teaching French." *Journal of Multilingual and Multicultural Development* 46 (10): 3643–3664. <https://doi.org/10.1080/01434632.2024.2355273>.

Gopalakrishnan, A. 2022. "Ecological Perspectives on Implementing Multilingual Pedagogies in Adult Foreign Language Classrooms – A Comparative Case Study." *International Journal of Multilingualism* 19 (1): 85–106. <https://doi.org/10.1080/14790718.2020.1712405>.

Gumperz, J. J. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511611834>.

Hasumi, T., and M.-S. Chiu. 2024. "Technology-Enhanced Language Learning in English Language Education: Performance Analysis, Core Publications, and Emerging Trends." *Cogent Education* 11 (1): 1–21. <https://doi.org/10.1080/2331186X.2024.2346044>.

Horwitz, E. K., M. B. Horwitz, and J. Cope. 1986. "Foreign Language Classroom Anxiety." *The Modern Language Journal* 70 (2): 125–132. <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>.

Hu, Y.-H., H.-Y. Yu, J.-W. Tzeng, and K.-C. Zhong. 2023. "Using an Avatar-Based Digital Collaboration Platform to Foster Ethical Education for University Students." *Computers & Education* 196:1–12. <https://doi.org/10.1016/j.compedu.2023.104728>.

Jiang, L., Z. Li, and J. S. C. Leung. 2024. "Digital Multimodal Composing as Translanguaging Assessment in CLIL Classrooms." *Learning and Instruction* 92:1–14. <https://doi.org/10.1016/j.learninstruc.2024.101900>.

Kaplan-Rakowski, R., and A. Gruber. 2023. "The Impact of High-Immersion Virtual Reality on Foreign Language Anxiety." *Smart Learning Environments* 10 (1): 1–18. <https://doi.org/10.1186/s40561-023-00263-9>.

Kousaie, S., and N. A. Phillips. 2012. "Ageing and Bilingualism: Absence of a "Bilingual Advantage" in Stroop Interference in a Nonimmigrant Sample." *Quarterly Journal of Experimental Psychology* 65 (2): 356–369. <https://doi.org/10.1080/17470218.2011.604788>.

Kress, G. 2009. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge. <https://doi.org/10.4324/9780203970034>.

Lai, C.-J. 2024. "Examining the Impact of Multimodal Task Design on English Oral Communicative Competence in Fourth-Grade Content-Language Integrated Social Studies: A Quasi-experimental Study." *Asian-Pacific Journal of Second and Foreign Language Education* 9 (1): 1–27. <https://doi.org/10.1186/s40862-024-00289-7>.

Lan, Y.-J. 2020. "Immersion into Virtual Reality for Language Learning." In *Psychology of Learning and Motivation*, edited by K. D. Federmeier and H.-W. Huang, 1–26. London: Academic Press. <https://doi.org/10.1016/bs.plm.2020.03.001>.

Lazovic, M. 2025. "Spatial Resources in Pre-service Teachers' Instructional Practices in VR Tandems: Co-constructing Shared Spaces and Embodied Spatial Scaffolding." *Frontiers in Communication* 10:1–25. <https://doi.org/10.3389/fcomm.2025.1519165>.

Lee, S.-M. 2023. "Second Language Learning through an Emergent Narrative in a Narrative-Rich Customizable Metaverse Platform." *IEEE Transactions on Learning Technologies* 16 (6): 1071–1081. <https://doi.org/10.1109/TLT.2023.3267563>.

Lee, S.-M., Z. Yang, and J. G. Wu. 2023. "Live, Play, and Learn: Language Learner Engagement in the Immersive VR Environment." *Education and Information Technologies* 29 (9): 10529–10550. <https://doi.org/10.1007/s10639-023-12215-4>.

MacIntyre, P., and T. Gregersen. 2012. "Affect: The Role of Language Anxiety and Other Emotions in Language Learning." In *Psychology for Language Learning: Insights from Research, Theory and Practice*, edited by S. Mercer, S. Ryan, and M. Williams, 103–118. London: Palgrave Macmillan UK. [https://doi.org/10.1057/9781137032829\\_8](https://doi.org/10.1057/9781137032829_8).

Marian, V. K., B. H. Elena, K. Ursula, and A.-K. Cordes. 2013. "Multilingual Stroop Performance: Effects of Trilingualism and Proficiency on Inhibitory Control." *International Journal of Multilingualism* 10 (1): 82–104. <https://doi.org/10.1080/14790718.2012.708037>.

Marre, Q., H. Nathalie, and E. Labeye. 2024. "Imagining Abstractness: The Role of Embodied Simulations and Language in Memory for Abstract Concepts." *Visual Cognition* 32 (1): 24–47. <https://doi.org/10.1080/13506285.2024.2375202>.

Mayer, R. E. 2024. "The Past, Present, and Future of the Cognitive Theory of Multimedia Learning." *Educational Psychology Review* 36 (1): 1–25. <https://doi.org/10.1007/s10648-023-09842-1>.

Myers-Scotton, C. 1993. *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198239055.001.0001>.

Noels, K. A., L. G. Pelletier, R. Clément, and R. J. Vallerand. 2000. "Why Are You Learning a Second Language? Motivational Orientations and Self-determination Theory." *Language Learning* 50 (1): 57–85. <https://doi.org/10.1111/0023-8333.00111>.

Parmani, A. 2023. "Virtual Reality in Language Learning: A Systematic Review and Implications for Research and Practice." *Interactive Learning Environments* 31 (1): 172–184. <https://doi.org/10.1080/10494820.2020.1765392>.

Pérez-Jorge, D., E. Olmos-Raya, A. I. González-Contreras, and I. Pérez-Pérez. 2025. "Technologies Applied to Education in the Learning of English as a Second Language." *Frontiers in Education* 10:1–13. <https://doi.org/10.3389/feduc.2025.1481708>.

Plonsky, L., and F. L. Oswald. 2014. "How Big Is "Big"? Interpreting Effect Sizes in L2 Research." *Language Learning* 64 (4): 878–912. <https://doi.org/10.1111/lang.12079>.

Poplack, S. 1980. "Sometimes I'll Start a Sentence in Spanish Y Termino En Español: Toward a Typology of Code-Switching." *Linguistics* 18 (7-8): 581–618. <https://doi.org/10.1515/ling.1980.18.7-8.581>.

Saint-Georges, I. d., and J.-J. Weber. 2013. *Multilingualism and Multimodality: Current Challenges for Educational Studies*. Rotterdam: Sense Publishers. <https://doi.org/10.1007/978-94-6209-266-2>.

Spechtenhauser, B., and U. Jessner. 2024. "Complex Interactions in the Multilingual Mind: Assessing Metalinguistic Abilities and Their Effects on Decoding a New Language System in Trilingual Learners." *Lingua. International Review of General Linguistics. Revue Internationale De Linguistique Generale* 301:1–24. <https://doi.org/10.1016/j.lingua.2024.103678>.

Steffensen, S. V., and C. Kramsch. 2017. "The Ecology of Second Language Acquisition and Socialization." In *Language Socialization*, edited by P. A. Duff and S. May, 1–16. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-02327-4\\_2-1](https://doi.org/10.1007/978-3-319-02327-4_2-1).

Tai, K. W. H., and L. Wei. 2024. "Mobilising Multilingual and Multimodal Resources for Facilitating Knowledge Construction: Implications for Researching Translanguaging and Multimodality in CLIL Classroom Context." *Journal of Multilingual and Multicultural Development* 46 (5): 1531–1541. <https://doi.org/10.1080/01434632.2024.2442525>.

Tang, F. 2024. "Understanding the Role of Digital Immersive Technology in Educating the Students of English Language: Does It Promote Critical Thinking and Self-directed Learning for Achieving Sustainability in Education with the Help of Teamwork?" *BMC Psychology* 12 (1): 1–14. <https://doi.org/10.1186/s40359-024-01636-6>.

Thrasher, T. 2022. "The Impact of Virtual Reality on L2 French Learners' Language Anxiety and Oral Comprehensibility." *CALICO Journal* 39 (2): 219–238. <https://doi.org/10.1558/cj.42198>.

Ukenova, A., G. Bekmanova, N. Zaki, M. Kikimbayev, and M. Altaibek. 2025. "Assessment and Improvement of Avatar-Based Learning System: From Linguistic Structure Alignment to Sentiment-Driven Expressions." *Sensors* 25 (6): 1–28. <https://doi.org/10.3390/s25061921>.

Valizadeh, M., and M. Morady Moghaddam. 2025. "Metaverse Magic: Improving Language Learners' Intercultural Understanding through Virtual Reality." *Journal of Multilingual and Multicultural Development* 46 (7): 2062–2081. <https://doi.org/10.1080/01434632.2025.2538691>.

Van Heuven, W. J., K. Conklin, E. L. Coderre, T. Guo, and T. Dijkstra. 2011. "The Influence of Cross-Language Similarity on Within- and Between-language Stroop Effects in Trilinguals." *Frontiers in Psychology* 2011:1–15. <https://doi.org/10.3389/fpsyg.2011.00374>.

Ware, A. T., M. Kirkovski, and J. A. G. Lum. 2020. "Meta-analysis Reveals a Bilingual Advantage That Is Dependent on Task and Age." *Frontiers in Psychology* 11:1–21. <https://doi.org/10.3389/fpsyg.2020.01458>.

Wei, L. 2018. "Translanguaging as a Practical Theory of Language." *Applied Linguistics* 39 (1): 9–30. <https://doi.org/10.1093/applin/amx039>.

Wong, G. K. W., and M. Notari. 2018. "Exploring Immersive Language Learning Using Virtual Reality." In *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, edited by M. J. Spector, B. B. Lockee, and M. D. Childress, 1–21. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-17727-4\\_144-1](https://doi.org/10.1007/978-3-319-17727-4_144-1).

Xie, Y., Y. Liu, F. Zhang, and P. Zhou. 2022. "Virtual Reality-integrated Immersion-based Teaching to English Language Learning Outcome." *Frontiers in Psychology* 12:1–10. <https://doi.org/10.3389/fpsyg.2021.767363>.

Yim, O., and R. Clément. 2021. "Acculturation and Attitudes toward Code-Switching: A Bidimensional Framework." *International Journal of Bilingualism* 25 (5): 1369–1388. <https://doi.org/10.1177/13670069211019466>.

York, J., K. Shibata, H. Tokutake, and H. Nakayama. 2021. "Effect of SCMC on Foreign Language Anxiety and Learning Experience: A Comparison of Voice, Video, and VR-Based Oral Interaction." *ReCALL* 33 (1): 49–70. <https://doi.org/10.1017/S0958344020000154>.

Žammit, J. 2023. "Exploring the Effectiveness of Virtual Reality in Teaching Maltese." *Computers & Education: X Reality* 3:1–11. <https://doi.org/10.1016/j.cexr.2023.100035>.

## Appendices

### Appendix A. Technical and pedagogical features of the multilingual metaverse platform

Component	Underlying Technology / API	Description	Pedagogical and Interactional Function
Platform Architecture	Unity-based metaverse application with RESTful API integration	Standalone installers for Windows, macOS, and Android connected to an institutional server via secure API endpoints	Ensured accessibility across devices and supported centralized data handling without requiring specialized hardware.
Input Modes	Unity microphone interface and in-game text chat module	Participants could alternate between spoken and typed input during interactions	Supported oral and written practice; allowed flexible engagement depending on learner preference and task type.
Output Modes	On-screen text renderer and text-to-speech (TTS) synthesis	NPC replies delivered in text and audio in target languages	Reinforced comprehension through multimodal exposure to linguistic input.
AI-Driven NPCs	OpenAI ChatGPT API (GPT-4) integrated via RESTful request handler	Generated multilingual, context-aware responses based on learner input and zone-specific parameters	Created authentic conversational dynamics and supported adaptive, code-sensitive dialogue.
Code-Switching Detection and Adaptation	Custom language identification module integrated with ChatGPT routing logic	Detected code-switching triggers (lexical, syntactic, or pragmatic) and adjusted NPC output accordingly	Promoted awareness of cross-linguistic boundaries and encouraged strategic language adaptation.
Zone-Language Mapping	Rule-based logic within Unity scene management	Mandarin = Chinatown; English = Campus; Filipino = Residential	Anchored language use to sociocultural and situational contexts reflective of authentic communicative environments.
Multimodal Interaction Features	Unity Animator controller and gesture scripts	Avatars performed gestures (wave, nod, shrug, questioning pose) automatically or manually	Reinforced meaning through embodied expression and multimodal signaling aligned with communicative intent.
Data Logging and Transmission	Secure API calls to institutional database (HTTPS protocol)	Interaction data (speech/text inputs, language switches, gesture activations, timestamps) sent in real time to the central analytics server	Enabled large-scale data capture for performance monitoring, linguistic analysis, and longitudinal study replication.

**Appendix B. Session-by-session breakdown of instructional activities, code-switching contexts, and learning outcomes across experimental and control conditions**

Session	Communicative Theme	Instructional Activities (Experimental)	Instructional Activities (Control)	Target Languages	Code-Switching Context	Learning Outcome
1	Greetings and Self-Introduction	Introduce self to NPCs; avatar gestures (wave, nod)	Scripted role-play: introductions in English and Filipino	English, Filipino	Switch to English for clarification or cultural references	Initiate conversations and perform self-introductions across languages
2	Making Simple Requests	Ask for assistance in various zones using voice/text input	Sentence drills and matching: common requests	English, Filipino	Switch to Filipino when assistance is misunderstood	Formulate and respond to basic requests using appropriate language
3	Shopping and Transactions	Interact with vendors in Chinatown using Mandarin	Vocabulary quiz and translation: shopping terms in Mandarin	Mandarin	Switch to English when vendor introduces unfamiliar item	Conduct transactional exchanges in Mandarin with limited support
4	Seeking Help and Giving Directions	Navigate campus to find locations using English and Filipino	Dialogue practice: giving and receiving directions	English, Filipino	Switch to English or Filipino when directions are unclear	Navigate and request information across zones using English and Filipino
5	Service Encounters (Formal Requests)	Simulate help desk tasks in campus area using formal English	Structured dialogue: service scenario in English	English	Switch to Filipino for informal clarification of formal request	Use formal English structures in service-related scenarios
6	Resolving Misunderstandings	Engage in role-play resolving conflict with NPCs across zones	Dialogue reconstruction: fixing a miscommunication	English, Mandarin	Switch to English to repair misunderstanding	Resolve communication breakdowns using repair strategies
7	Casual Social Interaction	Interact with dorm NPCs in Filipino; express opinions	Free conversation in pairs: weekend plans (Filipino)	Filipino	Switch to English when sharing personal opinion during disagreement	Express preferences and negotiate meaning in informal settings
8	Collaborative Task Completion	Complete group task requiring coordination across zones and languages	Written task: planning an event across languages	English, Filipino, Mandarin	Mixed code-switching based on task complexity and interlocutor	Collaborate in multilingual tasks requiring strategic language use

## Appendix C. Performance assessment rubrics

This appendix presents the rubrics used for the two performance-based components of the study: (C1) Communicative Competence and (C2) Code-Switching Proficiency.

Each dimension was rated on a five-point scale (1 = Emerging, 2 = Developing, 3 = Adequate, 4 = Competent, 5 = Proficient).

### C1. Communicative competence rubric

This rubric assesses multilingual communicative performance in video-recorded role-plays within the metaverse environment. It includes three interrelated dimensions that reflect linguistic, sociolinguistic, and pragmatic aspects of communication.

Dimension	Description	Performance Indicators (1–5 scale)
1. Linguistic Appropriateness	Accuracy and context-sensitive use of lexical, grammatical, and phonological forms across languages.	5: Consistently accurate and appropriate across languages; minor slips only. 4: Mostly accurate; minor errors that do not affect meaning. 3: Generally accurate with occasional breakdowns. 2: Frequent grammatical or lexical errors that reduce clarity. 1: Persistent inaccuracies that obscure intended meaning.
2. Language Choice	Effectiveness and appropriateness in selecting the language or code according to task, interlocutor, and context.	5: Language choice and switching are contextually appropriate and enhance communication. 4: Generally appropriate with a few inconsistent choices. 3: Some mismatches between code and context. 2: Recurrent inappropriate or random switches. 1: Language choice frequently disrupts communication.
3. Pragmatic Effectiveness	Ability to use language to achieve communicative goals in culturally and socially appropriate ways (e.g. making requests, refusals, clarifications).	5: Fully appropriate pragmatic choices; demonstrates cultural awareness. 4: Mostly appropriate with occasional pragmatic lapses. 3: Some awkward or unclear expressions but meaning conveyed. 2: Frequent pragmatic mismatches or inappropriate tone. 1: Pragmatic failures causing communication breakdowns.



## C2. Code-switching proficiency rubric

This rubric evaluates participants' ability to alternate languages appropriately and effectively during timed role-play scenarios.

Dimension	Description	Key Performance Indicators
1. Functional Appropriateness	Contextual relevance of language switching to communicative intent and task requirements.	5 – Proficient: Switches are purposeful and enhance communication. 4: Mostly appropriate; a few unnecessary or awkward switches. 3: Appropriate in general but inconsistent. 2: Occasional inappropriate or missing switches. 1: Frequent misuse or failure to switch when expected.
2. Linguistic Accuracy at Switch Points	Grammatical and lexical control during transitions between languages.	5: Smooth, accurate transitions; no interference errors. 4: Minor inaccuracies not affecting comprehension. 3: Occasional grammatical or lexical errors at switch points. 2: Frequent inaccuracies that hinder understanding. 1: Breakdowns at switch points causing loss of meaning.
3. Responsiveness and Fluency	Speed and naturalness of switching under communicative pressure.	5: Immediate, natural switching; fluent delivery. 4: Mostly fluent with brief hesitation. 3: Some hesitation or delayed switching. 2: Noticeable pauses or forced switching. 1: Hesitant, disfluent, or inappropriate switching.



## Appendix D. Representative learner-NPC interactions across the three language zones

Zone / Context	Speaker	Utterance	Interpretation
Chinatown District (Mandarin Zone)	Learner (L5)	'Zhège ... ano ... how much this one?'	Learner initiates purchase using Mandarin demonstrative ( <i>zhège</i> ) but shifts to Filipino filler and English due to lexical uncertainty, reflecting hybrid multilingual processing.
	NPC	'This one is thirty yuan. You can also say 'zhège duōshǎo qián?' in Mandarin.'	NPC provides corrective feedback by modeling full Mandarin phrase while ensuring comprehension through English translation.
	Learner (L5)	'Ah okay, zhège duōshǎo qián ... oh, thank you, xièxie!'	Learner repeats modeled structure and alternates languages pragmatically to confirm understanding.
	NPC	'Bù kèqi. Have a good day!'	NPC mirrors the learner's hybrid language use, reinforcing positive closure through code-switching.
Campus Area (English Zone)	Learner (L8)	'Good afternoon! I need to, ano po, extend my tuition fee payment ... pwede po ba?'	Learner opens formally in English but inserts Filipino ( <i>ano po, pwede po ba</i> ) for politeness and emotional nuance.
	NPC (Admin Officer)	'You may extend until Friday. Next time, you can say 'May I request an extension?''	NPC responds in English while modeling appropriate academic register, highlighting pragmatic scaffolding.
	Learner (L8)	'Ah okay po, thank you! I will remember, po.'	Learner blends English with Filipino honorific particle <i>po</i> , demonstrating culturally grounded politeness transfer.
Residential Area (Filipino Zone)	Learner (L2)	'Uy, pre! Did you finish the assignment in programming? Grabe, ang hirap naman nun!'	Informal peer interaction mixing Filipino and English; code-switching reflects topic-driven linguistic blending.
	NPC (Roommate)	'Oo, tapos na. Pero may Mandarin quiz tomorrow!'	NPC continues in Filipino with Mandarin lexical insertion ('Mandarin quiz'), showing natural multilingual mixing.
	Learner (L2)	'Ay, naku! Wō bù zhīdào, haha, I'm still noob sa Mandarin!'	Learner humorously inserts a simple Mandarin phrase ('Wō bù zhīdào' – I don't know) followed by English – Filipino code-mixing; this playful metalinguistic performance conveys camaraderie and emergent confidence rather than communicative necessity.