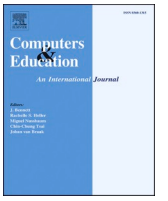
Computers & Education 248 (2026) 105593

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu



Investigating L2 listening comprehension and experience in immersive and interactive virtual reality: An experimental study

Yanting Liang [a,b], Guangwei Hu [c,*]

[a] Faculty of Arts and Sciences, Beijing Normal University, Zhuhai, China
[b] Faculty of Humanities, The Hong Kong Polytechnic University, China
[c] Department of English and Communication, The Hong Kong Polytechnic University, China

ARTICLE INFO

Keywords:
Virtual reality (VR)
Second language (L2) listening
Flow experience
Listening experience
L2 listening comprehension
EFL learners

ABSTRACT

Listening in the real world involves both verbal and non-verbal inputs. However, second language (L2) listening activities in the classroom often lack non-verbal inputs and are removed from the situational and cultural contexts where they would naturally occur. Virtual reality (VR) technology offers the potential to create more authentic and engaging L2 listening experiences. This study examines the impact of immersive and interactive VR on L2 listening experiences (operationalized as flow) and comprehension among Chinese university-level English-as-a-foreign-language (EFL) learners. Drawing on a randomized experimental design and semi-structured interviews, the study found that while VR did not directly improve L2 listening comprehension, it contributed indirectly to L2 listening comprehension by enhancing learners' listening experiences. Furthermore, although VR enhanced listening experiences in both cognitive and affective terms, only the affective enhancement had a noticeable positive medium-sized effect on L2 listening comprehension. Cognitive benefits of VR, such as sustained concentration and heightened absorption, did not translate into better L2 listening comprehension. The observed relationships can be explained by the misalignment between VR's interactive elements and the cognitive demands of the listening task. The findings highlight the pedagogical value of VR in enhancing affective engagement in learning, underscore the need for instructional design to mitigate cognitive overload, and emphasize the importance of careful VR design to ensure that immersive features support, rather than distract from, cognitive engagement.

1. Introduction

Listening, as a foundational language skill, is a dynamic cognitive activity that involves both verbal and non-verbal inputs. Non-verbal inputs, such as contextual features, gestures, eye contact, and facial expressions, provide important extralinguistic information and facilitate listening comprehension (Dahl & Ludvigsen, 2014; Sueyoshi & Hardison, 2005). However, second language (L2) listening activities, especially those in traditional classroom settings or using audio-based resources, often lack non-verbal inputs and are removed from the situational and cultural contexts in which they naturally occur. In addition to non-verbal inputs, the emotions experienced by L2 listeners also play a significant role in shaping listening outcomes. Prior research has shown that positive emotions such as foreign language enjoyment (FLE) have been linked to higher L2 achievement and engagement (Dewaele & Alfawzan, 2018; Li

* Corresponding author.
E-mail address: guangwei.hu@polyu.edu.hk (G. Hu).

https://doi.org/10.1016/j.compedu.2026.105593
Received 11 October 2025; Received in revised form 4 February 2026; Accepted 7 February 2026
Available online 9 February 2026

et al., 2019; Yang, 2021). Additionally, positive psychological interventions have been found to benefit English-as-a-foreign-language (EFL) learners' listening comprehension (Abdolrezapour & Ghanbari, 2021). Despite these findings, the impact of positive emotional experiences remains relatively underexplored, particularly in comparison with negative affect. While learning can elicit a wide range of emotions, such as anxiety, boredom, sadness, enjoyment, pride, and relief (Pekrun, 2014), anxiety has been frequently identified as a predominant negative emotion experienced by L2 listeners (Abdolrezapour & Ghanbari, 2021; Wang & MacIntyre, 2021), and has been shown to negatively impact listening comprehension (Kim, 2000; Mahmoodi et al., 2024; Zhang, 2013).

Recent developments in virtual reality (VR) technology offer new possibilities for addressing these problems by contextualizing L2 listening with more extralinguistic information and immersive interaction, and fostering more engaging listening experiences. In particular, immersive and interactive VR, which more fully leverages the medium's affordances, may hold greater potential to heighten learners' sense of presence and promote active involvement during listening tasks. In such environments, flow emerges as a particularly relevant positive psychological construct at the intersection of immersive VR experiences and language learning. It is defined as an optimal experience characterized by concentration, intense engagement, and intrinsic motivation (Csikszentmihalyi, 2000). Prior studies have demonstrated that flow is frequently elicited in technology-mediated language learning environments and plays a facilitative role in language learning performance (Karimi & Nasouri, 2024; Li et al., 2021; Liu et al., 2022; Wang & Feng, 2025). However, research on flow in L2 listening, especially in VR-based contexts, remains limited. As such, the present study seeks to address the issue of insufficient contextual support in traditional listening contexts, limited empirical research on positive emotional experiences in L2 listening, and the underexplored role of flow in VR-assisted language learning. Specifically, it aims to investigate how immersive and interactive VR influences L2 listening experiences and comprehension by addressing the following research questions.

RQ1. Does listening mode (VR vs. audio-only) influence L2 listening experiences?
RQ2. Do listening experiences influence L2 listening comprehension?
RQ3. Do listening experiences mediate the relationship between listening mode and L2 listening comprehension?

2. Literature review

2.1. VR and foreign language learning

A small body of empirical research has investigated VR's contributions to the development of foreign language skills. Wu et al. (2022), for example, examined the effect of VR on English-speaking performance, willingness to communicate (WTC) in English, and learning autonomy among 56 Taiwanese elementary students during a ten-week museum tour. The VR group, using VR software and head-mounted displays (HMD), significantly outperformed the control group (using slides and printed materials) in grammar and lexical use, but not in pronunciation, fluency, WTC, or autonomy. However, it is unclear whether the two groups were equivalent in these speaking dimensions prior to the intervention, as the pretest compared only overall speaking scores, making it difficult to attribute post-test dimension-level differences to the intervention. Adopting a mixed-methods research design, Ebadijalal and Yousofi (2024) assessed the impact of VR on writing motivation and performance among 42 EFL learners over nine weeks. The VR group, who took a virtual trip via Google Expeditions before writing, significantly outperformed the control group, who viewed pictures on tablets, in both writing motivation and writing performance. Similar to Wu et al. (2022), the study employed VR content that was largely observational with limited interactivity, which left one of VR's defining features underexplored.

As for listening, Tai and Chen (2021) randomly assigned 72 eighth-grade EFL students in Taiwan to either a VR or a video condition. In the VR condition, students played a language learning VR app using HMDs, while in the video condition, students watched a walkthrough video of the same content on a computer screen to complete five listening activities. The experimental study found that the VR group outperformed the video group in L2 listening comprehension and retention. However, it should be noted that the VR sessions incorporated speaking opportunities and script reference, which may have provided additional linguistic support beyond listening alone, thereby complicating the attribution of the observed advantages solely to the VR medium.

Not all studies have reported positive effects of VR on L2 listening comprehension. For example, Pinto et al. (2019) compared knowledge retention, sense of presence, and satisfaction in 12 EFL participants who sequentially engaged in VR and audio listening. The results suggested that while sense of presence and satisfaction were higher in VR listening, knowledge retention scores did not show a significant difference. However, this study did not report any reliability or validity evidence for the listening (knowledge retention) test and involved a very small sample size, which could undermine the robustness and interpretability of the findings. In contrast, Lee (2019) found that the VR group achieved significantly higher outcomes than the audio group and performed comparably to the video group in an L2 academic listening proficiency test. Similar to Pinto et al. (2019), this study also did not report reliability evidence for the listening performance measure, and the small sample size per condition ($n = 9$) could constrain statistical power.

Existing studies on VR-assisted language learning remain rather limited and often produce mixed findings. This pattern is corroborated by Dhimolea et al.'s (2022) systematic review of high-immersion VR for language learning. Prior research has suggested that the inconsistent effectiveness of VR could arise from its immersive and engaging properties, which may overload and distract learners, resulting in poorer learning performance (Makransky et al., 2019; Parong & Mayer, 2018). This pattern mirrors findings from research on computer-based educational games, which show that engaging game features can function as seductive details, imposing extraneous cognitive load and undermining learning when not carefully aligned with instructional objectives (Mayer, 2014). In light of the inconsistent empirical findings and methodological limitations in prior VR-assisted language learning studies, the present study seeks to provide a more rigorously controlled examination by addressing the lack of baseline comparability, under-exploration of VR interactivity, small sample sizes, and concerns regarding measurement validity.

2.2. Flow experience and foreign language learning

As a cognitive and affective construct relevant to technology-mediated learning experiences, flow has been widely examined in performance- and technology-mediated contexts, including VR. Flow typically occurs when certain preconditions (e.g., a balance between challenge and skill, unambiguous feedback, and clear goals) are met, resulting in deep concentration, a merging of action and awareness, loss of self-consciousness, a sense of control, altered perception of time, and an intrinsically rewarding (autotelic) experience (Csikszentmihalyi, 1991, 1993). It is important to consider the complexity of flow experience, both as a state and a trait. State flow refers to a transient optimal experience related to a specific activity, while trait flow reflects an individual's disposition to experience flow (Chirico et al., 2015). Prior research (Jackson & Eklund, 2002; Jackson et al., 1998; Jackson & Marsh, 1996) has shown that the intensity of flow can be reliably assessed with both state and dispositional flow scales. These two types of scales measure the same dimensions of flow and only differ in the wording of the instructions, asking respondents to reflect on either a particular situation (state flow) or general experiences (trait flow) (Jackson & Eklund, 2002).

The relationship between flow and learning can be understood through the General Model of Flow and Learning (Egbert, 2003), which distinguishes two groups of external antecedents of flow: contextual variables (e.g., learning environment, learning task, materials, and instructor's role) and learner characteristics (e.g., individual ability, prior experience, interest, personality, and learning style). This model positions flow as a mediator between the external antecedents and learning outcomes. To capture flow experience in educational settings, Heutte et al. (2021) postulated a refined four-dimensional flow model and developed the Education Flow Scale version 2 (for more details, see the section on instruments). They found that, in educational settings, not all original flow dimensions (Csikszentmihalyi, 1991, 1993) were consistently observed. Consequently, they restructured the model into four core dimensions. The first three dimensions – cognitive control, immersion and time transformation, and loss of self-consciousness – collectively measure the cognitive component (i.e., absorption) of flow experience, while the fourth dimension – autotelic experience – measures the affective component (see Fig. 1).

Previous research on flow experience in foreign language learning has suggested a positive effect of flow on language learning outcomes (Karimi & Nasouri, 2024; Li et al., 2021; Liu et al., 2022; Liu & Song, 2021; Wang & Feng, 2025). For example, Wang and Feng (2025) validated an existing flow model across different digital game-based vocabulary learning (DGBVL) apps. The study found that flow antecedents (e.g., skill-challenge balance and clear goals) significantly predicted flow experience (i.e., concentration, intrinsic motivation, and enjoyment), which in turn influenced outcomes (e.g., perceived learning and satisfaction). However, the study did not control for learners' dispositional tendency to experience flow (i.e., trait flow) and relied on perceived learning rather than objective performance measures to assess learning outcomes. Liu et al. (2022) monitored 216 Chinese EFL college students' flow experience in a semester-long writing course and found that a majority of them (66.4%) experienced flow with varying frequency. Flow was significantly correlated with intrinsic motivation and attention control, and higher flow frequency predicted better writing performance. Nevertheless, the interpretability of the results could be affected by the absence of validation for the adapted scales and possible inconsistencies in writing performance assessment across different instructors. Karimi and Nasouri (2024) distinguished between language- and game-induced flow in a study on the effects of digital text-based games (DTGs) on flow experience and incidental vocabulary learning. The study was innovative in distinguishing between game-induced and language-induced flow, showing that only the latter significantly predicted incidental vocabulary learning outcomes on both immediate and delayed post-tests. However, the relatively small sample size ($n = 57$) for the MANCOVA design with multiple dependent variables and covariates warrants cautious interpretation of the findings.

Extant research also suggests that technology-mediated language environments have the potential to foster flow experience (Egbert, 2003; Karimi & Nasouri, 2024; Li et al., 2021; Wang & Feng, 2025). However, most of these studies focused on conventional technology tools, such as mobile phone apps and online learning platforms. A notable gap remains in the literature regarding flow and L2 learning in multisensory, immersive, and interactive environments (e.g., VR). In addition, although the General Model of Flow and Learning (Egbert, 2003) positions flow as a mediating factor between external antecedents and learning outcomes, this relationship has not yet been empirically tested in VR contexts.

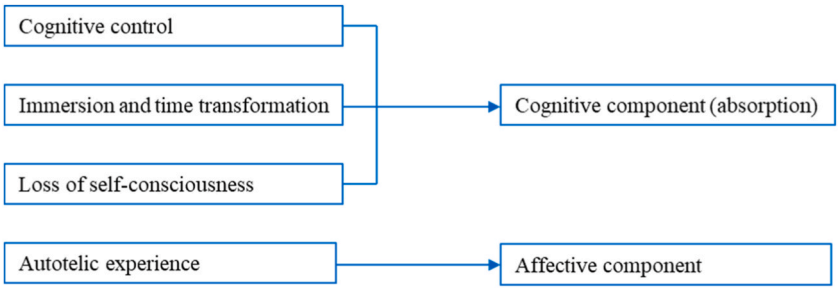


Fig. 1. The four dimensions of flow by Heutte et al. (2021).

2.3. Flow and performance in VR

Previous research has shown that a VR environment, with its interactive, immersive, and engaging features, facilitates flow experience. For example, Doğan et al. (2024) investigated the impact of immersive VR on flow experience in clinical law education among 83 law students, assigned to either an immersive VR or a desktop group. The findings showed that the VR group experienced significantly higher flow than the desktop group on four flow dimensions (i.e., attention, telepresence, time distortion, and interaction). Extant research, however, has produced mixed findings about the impact of flow on performance in VR. Tai et al. (2022) found that flow experience significantly predicted procedural accuracy and performance quality in a VR-based car-cleaning training program. Tai et al. (2024) reported that in VR drum practice, flow experience enhanced perceived learning value, which in turn predicted better rhythm performance. In a VR-based stargazing study, Tai and Hong (2025) found that higher learning interest and lower frustration enhanced flow experience in VR, which in turn positively predicted learning outcomes. By contrast, Keller et al. (2025) found that although their VR group significantly outperformed the desktop group in vocational training, flow experience had no significant effect on performance. Similarly, Yoon (2025) observed that a high level of flow did not necessarily translate into better performance. Her study compared the effects of VR simulation and small group practice (8–9 students per instructor) on peripheral intravenous catheterization (PIVC) skills, flow state, and learning satisfaction among 158 fourth-year nursing undergraduates. The results indicated that although the VR group reported significantly higher flow experience, the small group practice participants significantly outperformed the VR group in skill performance, due to immediate instructor feedback and hands-on practice.

Bian et al. (2018) found that the relationship between flow and performance in VR was often weak. They attributed this weak flow-performance link to incongruence between the interactive VR elements and the task to perform. Such incongruence may occur when VR interactions are distracting, unintuitive, or functionally misaligned with the primary task. Bian et al. (2018) showed that by improving the congruence between the VR interaction and the primary task in a VR tennis game, the flow-performance link became stronger. The issue of incongruence may result in increased extraneous processing, namely, cognitive processing that is not related to the instructional objective, and is influenced by the degree of distraction caused by immersive technologies (Mayer, 2001, 2022; Mayer et al., 2023). Due to the learner's limited cognitive capacity, increased extraneous processing may consume cognitive resources that would otherwise be available for cognitive processing that is important for meaningful learning, specifically, essential processing (required to mentally represent core learning content) and generative processing (devoted to organizing and integrating new knowledge into long-term memory and driven by the learner's motivation to learn) (Mayer et al., 2023).

To sum up, although VR has been generally found to induce flow experience, the effects of flow on performance are often inconsistent and weak, likely due to incongruence between interactive VR elements and the primary task. This incongruence may cause extraneous processing, distracting learners from task performance (Bian et al., 2018; Mayer et al., 2023). Notably, few studies have examined the flow–performance link in VR-based L2 learning, particularly from a comparative perspective.

3. Methods

3.1. Research design

To address the three research questions presented earlier, we designed a mixed-methods study that leveraged both quantitative and qualitative data to triangulate and complement our findings, thereby achieving a more comprehensive understanding (Creswell & Plano Clark, 2018; Green, 2007). Specifically, a sequential explanatory research design was adopted that comprised an initial quantitative strand and a subsequent qualitative strand. The quantitative strand was a randomized controlled trial, where 84 participants were randomly assigned to either a VR group or an audio group. The qualitative strand drew on semi-structured interviews with 21 participants from both groups to substantiate, illuminate, and complement the quantitative findings.

3.2. Participants

The 84 participants in this study were Chinese university EFL students who were aged between 18 and 25. Female students ($n = 72$) accounted for 85.7% of the sample, reflecting the gender distribution of the available student population. The participants came from 20 different majors, with English majors forming the largest subgroup and accounting for 35.7% ($n = 30$) of the sample. Based on self-reported prior examination attainment, the sample primarily comprised learners with intermediate to high levels of English proficiency: 71.4% ($n = 60$) reported having passed the College English Test Band 4 or Band 6 (CET-4/6), and 12.0% ($n = 10$) reported having passed the Test for English Majors Band 4 or Band 8 (TEM-4/8). The vast majority (83.3%) of the participants had little or no prior experience with VR. Using computer-generated random numbers, they were randomly assigned to either a VR group ($n = 42$) or an audio group ($n = 42$). The two groups did not differ in their pre-experiment listening proficiency measured by the listening subtest of the Cambridge B2 First for Schools ($t[82] = 0.88$, $p = .384$) or their trait flow scores ($t[82) = 0.04$, $p = .970$], suggesting comparable listening ability and general propensity to experience flow before the experiment.

3.3. Instruments

3.3.1. The EduFlow-2 scale

In this study, listening experience was operationalized as flow experience and measured using the EduFlow-2 scale developed by Heutte et al. (2021; see its appendix for the scale items). The same scale was used to assess participants' trait flow (i.e., general flow

experience across English learning activities as a disposition) before the experiment and state flow (i.e., situational flow experience) during the experiment. The only difference between the two administrations of the scale was in the instructions: the trait version asked participants to recall their general flow experience in English learning, whereas the state version prompted them to report the flow they were experiencing during the experiment. The EduFlow-2 scale measures four dimensions of flow (see Fig. 1) using 12 statements rated on a seven-point Likert scale (1 = strongly disagree, 7 = strongly agree), with three items assessing each dimension. The scale has been validated in educational contexts and shown to have strong psychometric properties, including good factorial validity, internal consistency, and construct validity (Heutte et al., 2021). In the present study, the scale's internal consistency was reassessed with SPSS 29.0, yielding a Cronbach's alpha of .90 for the trait version and .91 for the state version. Although alpha values approaching .90 warrant cautious interpretation with respect to potential item redundancy (Tavakol & Dennick, 2011), confirmatory factor analyses (CFAs) using IBM SPSS Amos 31.0 provided no evidence of item redundancy. Both state and trait flow scales demonstrated good model fit in the present sample (state flow: χ^2 (48) = 66.86, χ^2/df = 1.39, CFI = .98, TLI = .97, RMSEA = .07; trait flow: χ^2 (48) = 57.83, χ^2/df = 1.21, CFI = .99, TLI = .98, RMSEA = .05). Standardized factor loadings across both scales ranged from approximately .60 to .99, and inter-factor correlations remained below the .85 threshold, supporting convergent and discriminant validity.

3.3.2. The listening proficiency test

Participants' listening proficiency was assessed before the experiment using the listening subtest of the Cambridge B2 First for Schools (Cambridge University Press & Assessment, 2025a). Aligned with the B2 level of the Common European Framework of Reference for Languages (Council of Europe, 2020), the test covers a range of listening situations, from casual conversations to formal discourse, and assesses listening for gist, detail, and inference, reflecting the level of English proficiency typically required for English-medium higher education. According to Cambridge University Press & Assessment (2025a: Exam Format), the subtest "requires being able to follow and understand a range of familiar spoken materials, such as news programmes, public announcements and other sources." Research reports from Cambridge consistently demonstrate high reliability and strong evidence of construct and criterion validity for the listening component (Cambridge University Press & Assessment, 2025b: Quality and accountability). The listening subtest consists of 30 multiple-choice questions and lasts 40 min. Its reliability was estimated using participants' item-level responses from the present study. Cronbach's alpha, calculated in SPSS 29.0, was .82, indicating good internal consistency.

3.3.3. The demographic questionnaire

A demographic questionnaire (see Appendix I) was developed to collect information about participants' names, gender, age, academic year, major, VR-using experience, self-reported English proficiency, and the English proficiency tests that they passed. The questionnaire, written in Chinese (i.e., participants' first language), was administered just before the experiment and took 2-3 min to complete.

3.3.4. Listening materials

The listening materials used in the experiment were taken from the free educational VR application *Anne Frank House VR* (Vertigo Games & Knucklehead, 2024). The application situates VR players in the secret annex where Anne and her family hid from Nazi persecution during World War II. It offers an immersive historical and educational museum experience, allowing users to virtually explore the secret annex while listening to English diary excerpts narrated by a native female actor speaking from Anne Frank's first-person perspective. The VR application was selected after evaluating multiple VR educational applications, as it offers a high level of immersion and interactivity, an appropriate level of language difficulty, and a clear structure that guides learners throughout the session. In addition, its narrated content contains abundant factual details and implied messages, making it well aligned with the comprehension testing goal of the present study. To produce a comparable audio version of the monologue for the audio group, a screen recording was made during the VR walkthrough using the headset's native recording function. This ensured that the listening content, ambient background sounds, narrator's voice quality, and speech intelligibility were similar across the VR and audio modes. Nonetheless, we recognized that the intelligibility of the audio signals might differ slightly between headphone- and loudspeaker-based settings, as the VR application applied binaural rendering, which is more realistic and can provide a better signal-to-noise ratio. This subtle difference could influence task difficulty and individual effort.

3.3.5. Listening comprehension test

A listening comprehension test was developed based on *Anne Frank House VR* (see Appendix II for a sample of the test questions). The test comprised a practice section (four questions) and eight testing sections (totaling 30 multiple-choice questions). The practice section familiarized participants with the test format and procedure, and allowed them to adjust the audio volume for comfortable listening. Each of the eight testing sections featured a room exploration in Anne Frank's house and assessed both micro listening skills (e.g., recognizing factual details) and macro skills (e.g., recognizing the communication functions of utterances and inferring implied meanings). This tailor-made test was validated in a pilot study involving 55 participants drawn from the same student population as the main study. The reliability statistic ($\alpha = .807$) obtained indicated good internal consistency, and concurrent validity was evidenced by a substantial correlation ($r = .712$, $p < .01$) between participants' listening comprehension test scores and their Cambridge B2 First listening subtest scores. The reliability of the listening comprehension test was re-assessed with data collected for the present study, and the result ($\alpha = .81$) again indicated good internal consistency.

3.3.6. Semi-structured interviews

To collect qualitative data for triangulating, complementing, and explaining the quantitative findings, an interview guide was

developed prior to the experiment, based on the study's research questions and relevant theoretical frameworks (see Appendix III for the English version of the interview questions). The interview guide was structured around the study's core constructs (i.e., listening mode, flow experience, and listening comprehension) with questions designed to examine their interrelationships. Specifically, the interview questions focused on the perceived impact of the listening modes on listening experience, the state of flow during the listening task (e.g., skill–challenge balance, goal clarity, immersion, engagement, sense of control, absorption, enjoyment, and satisfaction), and the influence of flow on listening comprehension. The content validity of the interview questions was theoretically grounded in Egbert's (2003) General Model of Flow and Learning and Heutte et al.'s (2021) four-dimensional flow framework. In addition, the interview questions were piloted with five students (not participants in this study) to improve clarity and construct relevance. Using this interview guide, the first author conducted interviews with the selected participants individually. During the interviews, she followed up on interviewee responses with further probing questions and explored new topics that arose in the process. A purposeful sampling strategy was used to select the interviewees based on changes in pre- and post-test flow scores and listening scores. Participants whose scores showed average changes (representing typical cases) and exceptional differences (representing special cases) were selected to enhance the breadth and depth of the investigation. Interviewees were chosen from both the VR and audio groups to facilitate comparative analysis.

3.4. Procedure

Fig. 2 outlines the research procedure. Before the study began, all participants received an information sheet detailing the study procedures and were invited to provide informed consent. They were then assigned randomly to the VR or the audio group and completed the demographic questionnaire online. Prior to the experiment, they were administered the EduFlow-2 scale to measure their trait flow and the Cambridge B2 First listening subtest to assess their English listening proficiency. To mitigate VR's novelty effect and minimize technical issues during the experiment, participants in the VR group received a 30-min individual VR orientation. During this orientation, the experimenter demonstrated the operation of the VR equipment, and participants practiced basic interactions using several neutral VR applications, including a brief, non-narrative introductory scene from the Anne Frank House that was not part of the experimental content. Audio was muted throughout the orientation, and participants were not informed of the scene's context or storyline.

When the experiment began, participants completed the listening task in either VR or audio mode. In the VR condition, participants used a VR head-mounted device (Oculus Quest 2) and two hand trackers to navigate the rooms and activate audio segments by clicking sequentially highlighted icons (see Fig. 3). The audio was delivered via the headset's built-in speakers. The VR application followed a fixed, linear sequence: a new icon only appeared after the previous audio segment was completed, ensuring that all participants progressed through the VR experience in the same predetermined order. The audio group, on the other hand, listened to the audio via a classroom loudspeaker system connected to a PC, which was pre-calibrated to ensure clear and consistent sound quality.

The experiment for both groups was conducted in a standard university classroom equipped with desks, chairs, and a projector. The

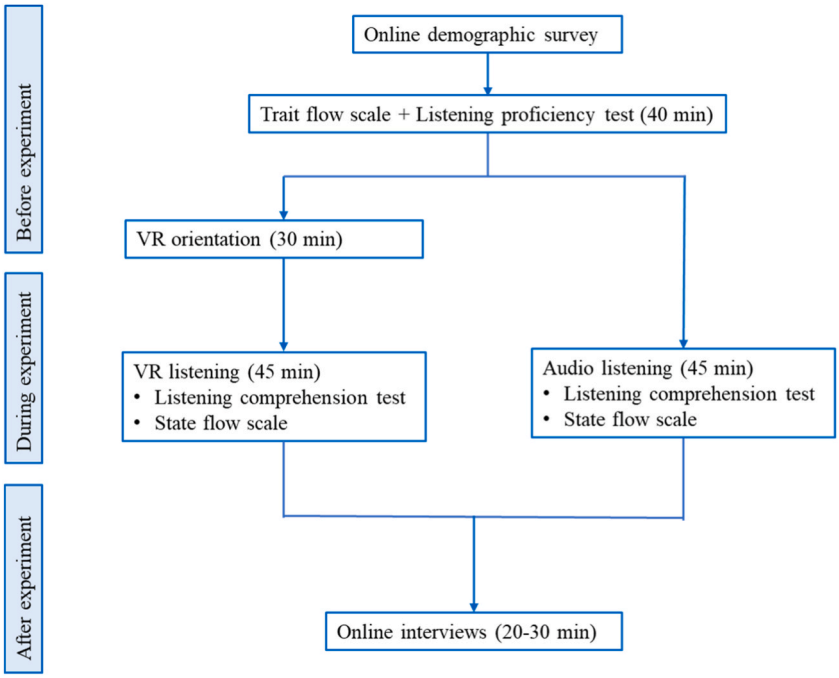


Fig. 2. Research procedures adopted in the study.

Fig. 3. Screenshots of the VR listening task.

audio group remained seated at their desks facing the front of the classroom during audio playback. For the VR group, desks and chairs were temporarily rearranged to create an open area that allowed safe movement during the VR session. Otherwise, the setting was identical to that of the audio condition. Prior to the listening task, both groups received a brief introduction through a PowerPoint slide that presented the basic setting and the names of key characters, without revealing any narrative details or information related to the test questions. This introduction ensured that both groups received equivalent and minimal contextual grounding.

As the listening task began, both groups completed a brief practice session to confirm that the volume was comfortable and clearly audible. No participants reported difficulty hearing the audio or, in the VR condition, viewing the content. The listening comprehension test accompanying the listening task was administered section by section. Participants had 10 s to read each testing question before a listening section and were allowed to answer the questions only after the corresponding audio was finished. For the VR group, the listening task was administered individually. To reduce memory load, the VR group removed their headset to answer the questions in each section. For the audio group, the listening comprehension test was administered in a group setting. They were not allowed to read or write while listening to match the VR group's task condition. Immediately after the listening task, both groups completed the EduFlow-2 scale again to measure their state flow based on the listening experience that they had just had.

Within 48 h of the experiment, individual semi-structured interviews were conducted online with 12 participants from the VR group and nine from the audio group. The interviews were conducted in Chinese and lasted on average 20 min. Participants responded orally during the interviews, which were video-recorded with their consent and transcribed verbatim for analysis.

3.5. Data analysis

To answer Research Questions 1 and 2 quantitatively, three hierarchical regression analyses were performed. The first two analyses examined whether listening mode predicted the cognitive and affective components of state flow, respectively (RQ1), whereas the third analysis examined whether the cognitive and affective components of state flow predicted listening comprehension (RQ2). Hierarchical regression analyses were chosen because they allowed for the assessment of the incremental effects of additional variables while controlling for those previously entered. The analyses were conducted with SPSS 29.0, with the alpha set at .05 (two-tailed). An *a priori* power analysis was conducted using G*Power 3.1.9.7 for a linear multiple regression model (fixed model, R^2 increase). Assuming a medium effect size ($f^2 = .15$) in line with conventional benchmarks for multiple regression in behavioral and educational research, an α level of .05, and a desired power of .80, the analysis indicated that a minimum sample size of 68 was required to test two predictors simultaneously. These predictors corresponded to the cognitive and affective components of flow, conceptualized as concurrent predictors of listening comprehension. As this model represents the most demanding analytic scenario, our obtained sample size of 84 was considered sufficient for all planned regression analyses involving fewer predictors.

To address Research Question 3, a mediation analysis was conducted using the PROCESS macro for SPSS (Hayes, 2022), which applies bootstrapping (5000 resamples) to generate bias-corrected 95% confidence intervals for robust estimates. The sample size ($N = 84$) was considered adequate for the planned mediation analysis. Simulation evidence suggests that, for mediation models with non-small indirect effects, adequate power (approximately .80) can normally be achieved with sample sizes below 100 using bootstrap methods (Fritz & MacKinnon, 2007).

To triangulate, illuminate, and expand the quantitative findings, a thematic analysis was conducted on the interview data using MAXQDA Analytics Pro 24. Braun and Clarke (2006) 6-phase procedure was followed for the analysis. First, the 21 interview transcripts were read repeatedly to achieve data familiarization and an overall understanding of participants' experiences. Second, initial codes were generated inductively through iterative coding, with segments of data labeled according to their relevance to the relationships among our key constructs. Third, related codes were collated to form candidate themes that captured recurring patterns. Fourth, these themes were reviewed and refined to ensure internal coherence and conceptual distinctiveness. Fifth, the finalized themes were defined to reflect their core meanings. Finally, the themes were illustrated with representative interview excerpts and interpreted in relation to the quantitative results. To ensure the trustworthiness of the qualitative analysis, peer debriefing and repeated discussions were employed throughout the analytic process to challenge initial interpretations, refine theme boundaries, and enhance analytic accuracy and conceptual clarity. In addition, pilot interview data were used to trial the initial coding approach, refine interview questions, and verify the clarity and feasibility of the analysis procedures prior to analyzing the main dataset. The results of

thematic analysis were summarized in a theme map (see Appendix IV) and integrated with the corresponding quantitative findings. The interview data were drawn from 21 participants, which was deemed sufficient to achieve thematic saturation, as prior methodological research has indicated that saturation in purposive qualitative interviews is typically reached within approximately 12 participants in relatively homogeneous samples (Guest et al., 2006).

4. Results

4.1. Factors influencing listening experiences

To address our first research question, we analyzed the experimental data and semi-structured interviews to understand the influence of listening proficiency, listening mode (i.e., VR vs. audio), and trait flow on listening experiences. We operationalized listening experiences as state flow, which includes a cognitive and an affective component (Heutte et al., 2021). The results are reported below according to the relevant outcome variables.

4.1.1. Factors influencing the cognitive component of state flow

The hierarchical regression analysis performed on the cognitive component of state flow included four predictors and their two-way interactions. Listening proficiency as an individual language skill was selected in view of the skill-challenge balance, which is a fundamental precondition to achieving flow (Csikszentmihalyi, 1991, 1993). The cognitive and affective components of trait flow were included because of their inherent connections with the corresponding components of state flow. Listening mode (i.e., VR vs. audio) was entered into the regression because previous research has shown that interactive and immersive VR can influence flow experience (Doğan et al., 2024). Interaction terms involving the four variables were added to explore whether they jointly shaped state flow. The predictors were entered in the following order.

Step 1: listening proficiency
Step 2: cognitive component of trait flow; affective component of trait flow
Step 3: listening mode
Step 4: listening mode × listening proficiency; listening mode × trait flow's cognitive component; listening mode × trait flow's affective component

The listening proficiency and trait flow scores at Step 4 were centered by subtracting the grand mean from each participant's score. All assumptions underlying hierarchical regression were met: linearity and homoscedasticity were visually assessed; the Shapiro–Wilk test confirmed approximate normality of residuals ($W = .98$, $p = .124$); the Durbin–Watson statistic indicated no autocorrelation (DW = 2.07); and all Variance Inflation Factor (VIF) values were below 5, suggesting no multicollinearity.

In Model 1 (see Table 1), listening proficiency significantly predicted the cognitive component of state flow, ($B = 0.43$, $\beta = .25$, $p < .05$), explaining 6% of its variance ($r^2_{partial} = .06$). Model 1 was statistically significant, adjusted $R^2 = .05$, $F(1, 82) = 5.54$, $p < .05$.

In Model 2, the cognitive component of trait flow significantly predicted the cognitive component of state flow ($B = 0.61$, $\beta = .54$, $p < .001$), explaining a unique 22% of its variance ($r^2_{partial} = .22$), whereas its affective component was not a significant predictor. Notably, listening proficiency was no longer a significant predictor ($B = 0.23$, $\beta = .13$, $p = .150$) once the two components of trait flow were included, indicating that its effect was largely accounted for by trait flow. Model 2 was statistically significant, adjusted $R^2 = .36$, $F(3, 80) = 16.57$, $p < .001$, explaining an additional 32% of the variance in the dependent variable, $\Delta R^2 = .32$, $\Delta F(2, 80) = 20.76$, $p < .001$.

Listening mode, entered in Model 3, significantly predicted the cognitive component of state flow ($B = 6.85$, $\beta = .36$, $p < .001$), explaining an additional 20% of its variance ($r^2_{partial} = .20$) beyond the effects of the previously entered predictors. This indicated that the VR listening mode significantly increased cognitive absorption during the listening task. The cognitive component of trait flow continued to significantly predict the cognitive component of state flow, while its affective component and listening proficiency were

Table 1
Results of the hierarchical regression analysis on the cognitive component of state flow ($N = 84$).

Variable	Model 1			Model 2			Model 3			Model 4		
	B	SE *B*	β	*B*	SE B	β	*B*	SE *B*	β	*B*	SE *B*	β
Listening proficiency	0.43	0.18	.25*	0.23	0.16	.13	0.17	0.14	.10	0.34	0.21	.20
Cognitive component of trait flow				0.61	0.13	.54***	0.54	0.12	.48***	0.69	0.16	.60***
Affective component of trait flow				0.15	0.26	.07	0.32	0.24	.14	0.25	0.36	.11
Listening mode							6.85	1.55	.36***	6.94	1.53	.36***
Listening mode × listening proficiency										−0.35	0.28	−.15
Listening mode × cognitive component										−0.35	0.24	−.23
Listening mode × affective component										0.31	0.48	.11
R^2			.06			.38			.51			.53
Adjusted R^2			.05			.36			.48			.49
F Change			5.54*			20.76***			19.49***			1.59

*$p < .05$; **$p < .01$; ***$p < .001$.

non-significant predictors. Model 3 explained 48% of the variance, adjusted $R^2 = .48$, $F(4, 79) = 20.18$, $p < .001$. The addition of listening mode significantly improved the model, $\Delta R^2 = .12$, $\Delta F(1,79) = 19.49$, $p < .001$.

None of the three interaction terms entered into Model 4 emerged as significant predictors. The cognitive component of trait flow and listening mode remained the only significant predictors. Model 4 explained 49% of the variance, adjusted $R^2 = .49$, $F(7, 76) = 12.47$, $p < .001$, and the addition of the interaction terms did not significantly improve the model, $\Delta R^2 = .03$, $\Delta F(3,76) = 1.59$, $p = .200$.

The interviews provided qualitative evidence (see Sub-theme 1 under Theme 1: Cognitive experience, Appendix IV) in support of the quantitative finding that VR significantly enhanced the cognitive aspects (e.g., immersion, concentration, and cognitive control) of listening experiences. Specifically, all the interviewees from the VR group reported that VR created a sense of immersion. One participant (SV7) described this as resulting from "richer sensory stimulation" and "tactile, visual, auditory effects." The combination of these sensory inputs created an authentic environment where participants felt as if they were physically present in a real-world context. For example, SV10 stated that "VR made me feel as if I were actually there—it provided a stronger sense of immersion, making it easier to get fully engaged in the environment."

Furthermore, more than half ($n = 7$) of the VR interviewees mentioned that VR enhanced their concentration by anchoring their attention to the unfolding events. As SV1 said, "with such concrete scenarios and specific tasks, I felt that my concentration significantly improved compared to the pre-test." More than half of the interviewees from the VR group ($n = 7$) reported a moderate to high level of sense of control as VR allowed them to actively navigate the listening task, develop agency over task execution, and benefit from the contextual cues afforded by VR. Such cues supplied more information, facilitated inference, and accelerated cognitive processing, thereby increasing their perceived understanding. As SV10 explained, "with other types of listening, you're quite passive – someone else plays the audio, and all you can do is listen. But with this one, you could control when it spoke …. Seeing the environment in VR, rather than just listening, gave a clearer context and made the speech easier to understand." In contrast, only two interviewees from the audio group (22%) reported a high sense of control, due to "the lack of visual support" (SA3 and SA8).

4.1.2. Factors influencing the affective component of state flow

A hierarchical regression analysis run on the affective component of state flow included the same predictors and their two-way interactions as reported in the preceding section, with the variables entered in the same order as in the previous analysis. All the key statistical assumptions were met: linearity and homoscedasticity were visually confirmed; the residuals were approximately normally distributed (Shapiro–Wilk test: $W = .98$, $p = .173$); independence of errors was supported by the Durbin–Watson statistic (DW = 1.91); and all VIF values were below 5, indicating no concerns about multicollinearity.

As shown in Table 2, listening proficiency in Model 1 was not a significant predictor of the outcome variable. The overall model was not statistically significant, adjusted $R^2 = .02$, $F(1, 82) = 2.43$, $p = .123$. Notably, in all subsequent models, listening proficiency remained a non-significant predictor, indicating that it played a minimal role in shaping learners' emotional engagement during the listening task.

In Model 2, the cognitive component of trait flow was not a significant predictor, but its affective component significantly predicted the affective component of state flow ($B = 0.34$, $\beta = .39$, $p < .01$), explaining a unique 10% of the variance ($r^2_{\text{partial}} = .10$). Model 2 was statistically significant, adjusted $R^2 = .11$, $F(3, 80) = 4.51$, $p < .01$, and explained an additional 12% of the variance beyond Model 1, $\Delta R^2 = .12$, $\Delta F(2, 80) = 5.41$, $p < .01$.

In Model 3, listening mode significantly predicted the affective component of state flow ($B = 3.33$, $\beta = .45$, $p < .001$), uniquely explaining 23% of its variance ($r^2_{\text{partial}} = .23$), indicating that VR significantly enhanced affective engagement during the listening task. The affective component of trait flow showed a consistent impact ($B = 0.43$, $\beta = .48$, $p < .001$, $r^2_{\text{partial}} = .18$). Model 3 explained 31% of the variance, adjusted $R^2 = .31$, $F(4, 79) = 10.20$, $p < .001$, showing a significant improvement over Model 2, $\Delta R^2 = .20$, $\Delta F(1,79) = 23.48$, $p < .001$.

In Model 4, none of the three newly entered interaction terms were significant predictors. As with the previous models, the affective component of trait flow and listening mode continued to show a significant impact. Model 4 explained 31% of the variance, adjusted $R^2 = .31$, $F(7, 76) = 6.21$, $p < .001$, and the addition of the interaction terms did not significantly improve the model, $\Delta R^2 = .02$, ΔF

Table 2
Results of the hierarchical regression analysis on the affective component of state flow ($N = 84$).

Variable	Model 1			Model 2			Model 3			Model 4		
	B	SE *B*	β	*B*	SE *B*	β	*B*	SE *B*	β	*B*	SE *B*	β
Listening proficiency	0.11	0.07	.17	0.10	0.07	.15	0.07	0.06	.11	0.17	0.10	.25
Cognitive component of trait flow				−0.04	0.06	−.09	−0.07	0.05	−.16	−0.12	0.07	−.27
Affective component of trait flow				0.34	0.12	.39**	0.43	0.10	.48***	0.57	0.16	.63***
Listening mode							3.33	0.69	.45***	3.34	0.69	.45***
Listening mode × listening proficiency										−0.16	0.13	−.18
Listening mode × cognitive component										0.08	0.11	.14
Listening mode × affective component										−0.23	0.22	−.21
R^2			.03			.14			.34			.36
Adjusted R^2			.02			.11			.31			.31
F Change			2.43			5.41**			23.48***			.92

*$p < .05$; **$p < .01$; ***$p < .001$.

(3,76) = .92, p = .434.

Findings from the interview data were consistent with the quantitative results reported above (see Sub-theme 2 under Theme 1: Affective experience, Appendix IV). All the interviewees from the VR group reported positive emotional influences of the VR listening mode, including intrinsic motivation and enjoyment, reduced task pressure, and greater empathy and connection. A great majority of the VR interviewees (83.3%) reported that the visually rich, authentic environment, together with interactive objects and sound effects, transformed the listening task into an enjoyable and engaging experience. As SV5 explained, "I approached the listening task with the mindset and enjoyment of playing a game. This made me more inclined and willing to engage in English listening." This intrinsic motivation contrasts with the audio group, where some participants felt "less engaged" (SA3) and reported relying primarily on extrinsic motivation, such as viewing the task as "an opportunity for listening practice for an upcoming exam" (SA5 and SA2).

One-fourth of the VR interviewees also confided that the VR listening experience reduced test-taking pressure. Some mentioned that the experience felt "less of a formal test" (SV8) and helped them break away from the "mechanical feeling often associated with test-taking" (SV2). Lastly, half of the VR interviewees reported that, compared to "the audio-only mode [they typically use during their regular listening practice], which felt rather cold and distant" (SV11), the VR listening mode fostered greater empathy and connection. VR allowed them to "empathize more easily" (SV6), "felt like being with Anne" (SV6) throughout the story, making the listening experience more personal and emotionally resonant. In contrast, the audio mode, as reported by one interviewee in the audio group, was perceived as "lacking personal relevance" (SA3).

4.2. Factors influencing listening comprehension

To address our second research question, a third hierarchical regression analysis was performed on listening comprehension. The predictors were listening proficiency, the cognitive component of state flow, the affective component of state flow, and their two-way interactions. They were entered in the following order.

Step 1: listening proficiency
Step 2: cognitive component of state flow; affective component of state flow
Step 3: listening proficiency × affective component of state flow; listening proficiency × cognitive component of state flow

The regression analysis met the assumptions of linearity and homoscedasticity. Independence of errors was supported by the Durbin-Watson statistic (DW = 2.06). Multicollinearity was not a concern (all VIFs <3). Although the Shapiro-Wilk test indicated non-normality (W = .96, p < .01), the Normal Q-Q plot showed only minor deviations at the tail. Given the robustness of regression analyses to slight normality violations and the adequate sample size, the analysis was deemed appropriate.

In Model 1 (see Table 3), listening proficiency significantly predicted listening comprehension (B = 0.62, β = .68, p < .001), uniquely accounting for 46% of its variance ($r^2_{partial}$ = .46). The model was statistically significant, adjusted R^2 = .45, F(1, 82) = 69.84, p < .001. In Model 2, the affective component of state flow showed a borderline significant yet medium-sized effect on listening comprehension (B = 0.27, β = .20, p = .051), accounting for a unique 5% of the variance ($r^2_{partial}$ = .05), whereas its cognitive component was not a significant predictor (B = 0.04, β = .08, p = .456). As with the previous model, listening proficiency remained the strongest predictor. Model 2 was statistically significant, adjusted R^2 = .50, F(3, 80) = 28.78, p < .001, and improved over Model 1, ΔR^2 = .06, ΔF(2, 80) = 4.92, p < .01. The results pointed to a role of affective engagement – rather than cognitive absorption – in supporting listening comprehension. In Model 3, neither of the two newly entered interaction terms was a significant predictor of listening comprehension. Listening proficiency remained a strong predictor (B = 0.58, β = .64, p < .001), whereas the affective and cognitive components of state flow became non-significant. Model 3 explained 50% of the variance, adjusted R^2 = .50, F(5, 78) = 17.76, p < .001, with no significant improvement over Model 2, ΔR^2 = .01, ΔF(2, 78) = 1.10, p = .337.

The quantitative finding that only the affective component of state flow, rather than its cognitive component, significantly contributed to listening comprehension was illuminated by the interview data (see Sub-theme 1 under Theme 1: Cognitive influences, Appendix IV). While some interviewees mentioned some positive cognitive effects, such as absorption, enhanced concentration, and increased sense of control, many reported that the abundance of VR contextual details induced cognitive distraction. Over half of the interviewees in the VR group mentioned that their attention was diverted away from the listening task by the rich contextual elements,

Table 3
Results of the hierarchical regression on listening comprehension (N = 84).

Variable	Model 1			Model 2			Model 3		
	B	SE B	β	B	SE B	β	B	SE B	β
Listening proficiency	0.62	0.07	.68***	0.57	0.07	.63***	0.58	0.07	.64***
Affective component of state flow				0.27	0.14	.20*	0.24	0.14	.17
Cognitive component of state flow				0.04	0.05	.08	0.05	0.05	.10
Listening proficiency × affective component							−0.03	0.03	−.14
Listening proficiency × cognitive component							0.01	0.01	.16
R^2			.46			.52			.53
Adjusted R^2			.45			.50			.50
F Change			69.84***			4.92**			1.10

*p < .05; **p < .01; ***p < .001.

which caused them to miss key details, and thus negatively affected their comprehension. As SV2 noted, "at first, my attention was all on the scene, and I didn't pay attention to what it was saying. As a result, I got three out of four test questions wrong." This viewpoint was also supported by one-third of the interviewees in the audio group, who mentioned that the pure audio mode allowed them to maintain "full engagement" and kept them in "a constant state of intense listening" (SA2).

Apart from the contextual elements of the VR condition, the intense listening experience also introduced cognitive distractions and, in some cases, disruptions during the listening task (see Sub-theme 2 under Theme 2: Negative influences, Appendix IV). In several interview accounts of intense flow, participants described instances of "passive listening" (SV6), in which attention was diverted away from active auditory processing (SV2, SV7, SA9). When transitioning out of a deep flow state, SA6 reported feeling momentarily disoriented and struggling to recall information and process what she had heard. She explained, "when I felt immersed, I sometimes had trouble clearly hearing what was being said. I was completely immersed in the scene – she was speaking, and I could hear the sound, but I wasn't actively trying to grasp the information. It was like my attention wasn't fully focused on the listening task anymore."

Taken together, the quantitative and qualitative results showed that while VR enhanced both cognitive and affective aspects of listening experiences, only the affective benefits had a positive, medium-sized effect on listening comprehension. In contrast, its cognitive benefits (e.g., increased concentration and absorption) did not enhance listening comprehension, possibly due to cognitive distraction or overload induced by the immersive and interactive VR context and the intensity of the listening experience.

4.3. The mediating role of listening experiences

To address the third research question (i.e., whether state flow mediated the relationship between listening mode and listening comprehension), a mediation analysis using PROCESS Model 4 (Hayes, 2022) was conducted with trait flow and listening proficiency as covariates. As can be seen from Table 4 and Fig. 4, listening mode significantly predicted state flow ($B = 9.96$, SE $= 1.98$, $p < .001$), which, in turn, significantly predicted listening comprehension ($B = 0.09$, SE $= 0.05$, $p = .046$). However, the direct effect of listening mode on listening comprehension was not significant ($B = 0.91$, SE $= 0.91$, $p = .323$). Notably, the indirect effect of listening mode on listening comprehension via state flow was significant, as the 95% bootstrap confidence interval did not include zero ($B = 0.91$, BootSE $= 0.39$, 95% CI [0.28, 1.79]), indicating that state flow fully mediated the relationship.

The interview data provided additional insights into how flow shaped listening comprehension (see Theme 2, Appendix IV). After being briefly introduced to the concept of flow, 75% of the interviewees in the VR group reported experiencing flow during the listening task, whereas less than half ($n = 4$) in the audio group reported such a state. Regardless of whether they experienced flow during the experiment, an overwhelming majority (85.7%) of the interviewees from the two groups perceived potential effects of flow on their listening comprehension. They reported that flow experience helped them achieve more "global" (SV2, SV5, and SV8) and "precise" comprehension (SV5, SV9, SV10, SA5, and SA8), along with effortless remembering of information (SV1, SV3, and SA6). As SV1 explained, "even if I forgot what I was supposed to listen for at the time, I was still able to recall the relevant information later when I looked at the test question. I could remember that I had heard the related information in a specific context" (SV1). SV9 observed that "I believe there's a positive relationship between flow and my performance: the more engaged I am, the better my listening performance tends to be." Similarly, SV8 shared that "it (flow) definitely had an impact. I felt that it made me focus more on the overall content. It mainly influenced how I absorbed information. What I was able to grasp when I was in a state of flow was clearly different from when I wasn't." However, SV4 and SV6 also noted cognitive disruptions resulting from the intense state.

To sum up, the findings reported above indicated that the effect of listening mode on listening comprehension was mediated by listening experience (i.e., state flow), which was found to support precise and global comprehension and memory retention despite occasional potential cognitive disruptions.

5. Discussion

5.1. State flow as a mediator between VR and listening comprehension

The present study revealed that VR positively influenced listening comprehension primarily through its enhancement of listening experiences, rather than through a direct effect on listening comprehension. The mediating role of listening experiences (i.e., state flow) between listening mode and listening comprehension is supported by the General Model of Flow and Learning (Egbert, 2003), which postulates flow as a mediator between external antecedents (e.g., VR) and learning outcomes (e.g., listening comprehension). This theoretical model also helps explain why the hierarchical regressions in the present study accounted for large, though not overwhelming, proportions of the variance in listening experience and comprehension. According to the theoretical model, external

Table 4
Mediation analysis of the effect of listening mode on listening comprehension via state flow.

Relationship	*B*	SE	*p*	95% Confidence Interval	
				Lower	Upper
Listening mode → State flow	9.96	1.98	<.001	6.01	13.90
State flow → Listening comprehension	0.09	0.05	.046	0.002	0.18
Listening mode → Listening comprehension (direct)	0.91	0.91	.323	−0.91	2.73
Listening mode → State flow → Listening comprehension (indirect)	0.91	0.39	—	0.28	1.79

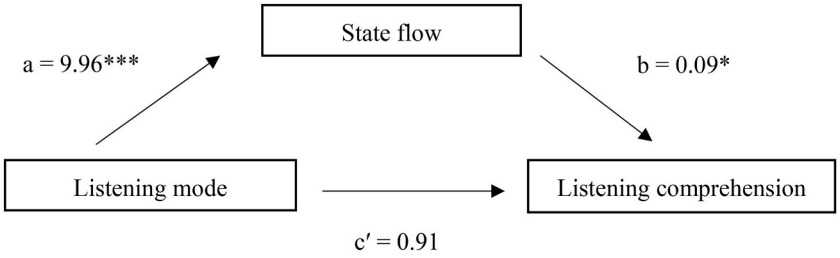


Fig. 4. Mediation model of listening mode, state flow, and listening comprehension.

antecedents of flow include both contextual factors (e.g., learning environment, instructional materials, and task design) and learner characteristics (e.g., personality, learning style, and prior experience). The learning medium itself, such as VR, represents only one element among these multiple factors. Among them, individual differences have been empirically shown to be an important contributor to flow experience. For example, Ellis et al. (1994) found that person-level factors accounted for up to approximately 20% of the variance in affective indicators of flow, over and above situational factors. As such, the hierarchical regressions that explained 49% and 31% of the variance in cognitive and affective flow experiences, respectively, and 50% of the variance in listening outcomes are consistent with theoretical expectations and empirical patterns in flow research.

The absence of a direct VR effect on listening comprehension could be explained by a widely accepted view in educational technology scholarship: media itself does not directly enhance learning (Clark, 1994). Instead, it is the design of learning experiences and the pedagogical strategies enabled by technology that matter most. From this perspective, VR does not inherently improve listening comprehension. Rather, it creates conditions, such as immersion, interactivity, and engagement, that foster a learning-conducive experience, which in turn enhances performance and learning.

Positive effects of flow on learning and performance have been reported by previous flow research on language learning in general contexts (e.g., Karimi & Nasouri, 2024; Li et al., 2021; Liu et al., 2022; Wang & Feng, 2025) and skill development/performance in VR contexts (e.g., Tai et al., 2022, 2024; Tai & Hong, 2025). Such positive effects of flow on learning and performance could be attributed to a neurocognitive basis. Dietrich (2004) postulates possible neurocognitive mechanisms underlying the experience of flow in terms of explicit and implicit information-processing systems. He notes that theoretical and empirical research in cognitive and neural sciences suggest that the explicit system is "associated with the higher cognitive functions of the frontal lobe and medial temporal lobe structures" and serves to enhance cognitive flexibility, whereas the implicit system is linked to "the skill-based knowledge supported primarily by the basal ganglia and has the advantage of being more efficient" (p.746). Drawing on research on sensory-motor integration skills, he proposes that flow occurs when skills that are highly practiced and encoded in the knowledge base of the implicit system are performed without any interference from the explicit system. This temporary suspension of the explicit system allows the implicit system, which operates automatically with minimal effort, to take over. The full mediation of the relationship between listening mode and listening comprehension by state flow, as found in the present study, is consistent with these postulated neurocognitive mechanisms underlying the flow experience. It suggests that when participants were in a flow state, they switched from conscious, effortful analysis to more automatic and intuitive processing, enabling them to concentrate better, retain more information, and ultimately perform better.

5.2. Immersed but distracted: the cognitive-affective paradox in VR

The present study found that VR enhanced listening experiences in both cognitive and affective terms. This finding aligns with previous research on VR-assisted language learning, which has shown that listeners generally exhibit a positive attitude toward the use of VR for language learning (Tai & Chen, 2021; Wu et al., 2022). While VR was found in this study to produce significant cognitive and affective enhancements in participants' state flow, only the affective enhancement had a marginally significant, medium-sized effect on listening comprehension. The role of the affective component of state flow (i.e., autotelic experience characterized by a sense of well-being and enjoyment) in facilitating learning performance is consistent with the Cognitive-Affective Theory of Learning with Multimedia (Moreno, 2005; Moreno & Mayer, 2007), which posits that affective and motivational factors play a critical role in promoting cognitive engagement, especially generative processing, or cognitive processing that is driven by the learner's motivation to actively make sense of the learning material. In contrast, the cognitive benefits of VR, such as heightened concentration and absorption, did not translate into better performance, likely due to distractions caused by VR's contextual elements, as revealed by the interviews. Specifically, in the context of the present study, not all the contextual elements in the *Anne Frank House VR* aligned closely with the auditory information. For example, some accessible corners and manipulable objects in the house were not directly linked to the listening content. While these immersive and interactive elements enhanced cognitive absorption, this heightened cognitive resource may have been directed toward VR explorations rather than the listening task.

In view of the findings discussed above, it can be concluded that the impact of VR's contextual elements depends on how well they align with the task goal. Well-aligned contextual elements closely correspond to the specific details in the listening material, anchoring participants' attention both on the task and the immersive environment, and making the immersive elements an integral part of the task. In contrast, misaligned elements often do not directly match the listening content, potentially causing cognitive dissonance or distractions. Thus, participants may still experience flow, but the focus of their flow experience may shift towards the immersive VR

environment rather than the task at hand. This interpretation is consistent with Bian et al.'s (2018) finding that when interactive elements in VR were misaligned or "disjoint" with the task, they could introduce distractions and lead to a flow experience detached from the task, resulting in a weak flow-performance link. This "incongruence" explanation also has a theoretical basis. According to the Cognitive Theory of Multimedia Learning (Mayer, 2001), irrelevant and distracting elements in immersive VR often cause increased extraneous processing (i.e., cognitive processing that does not serve the task-performing goal). Due to limited cognitive processing capacity, the increased extraneous processing can consume cognitive resources that would otherwise be available for meaningful cognitive processes, such as essential processing (mentally representing the learning content) and generative processing (organizing and integrating the new information with prior knowledge to make sense of the learning content) (Mayer et al., 2023). As such, the cognitive benefits of VR may depend on both how well the contextual elements align with the task and how effectively participants can regulate their attention when faced with disjoint or misaligned contextual elements.

Apart from cognitive distraction, another issue that may negatively affect VR-assisted learning outcomes is the cognitive overload induced by rich multisensory stimulation. Nevertheless, prior research has suggested that such overload can be reduced through instructional design without undermining engagement, for instance, by segmenting VR lessons to better regulate essential processing demands and supplementing immersive exposure with brief generative activities, such as summarization tasks between segments (Parong & Mayer, 2018). Another way to mitigate cognitive overload is through careful VR design that aligns immersive elements with the target cognitive task, thereby reducing extraneous cognitive load.

Although the present findings indicate that VR's cognitive affordances did not translate into immediate performance gains, fully understanding the nuances of VR's cognitive impact requires moving beyond outcome measures to consider its influence on learners' cognitive processing more directly. While it is unlikely that VR could alter stable cognitive capacities such as working memory or attentional control over a short period of time, it may prompt situational adaptation in how learners deploy these capacities during task engagement. From this perspective, VR's cognitive influence could be better understood as situational and process-oriented rather than capacity-enhancing. This perspective may help explain why significant experiential effects (more state-dependent) were observed alongside limited cognitive performance outcomes in the present study.

6. Conclusion

This study examined the impact of immersive and interactive VR on L2 listening experiences (operationalized as flow) and comprehension among 84 Chinese university-level EFL learners. Drawing on a randomized experimental design and semi-structured interviews, the study found that while VR did not directly improve L2 listening comprehension, it contributed indirectly to L2 listening comprehension by enhancing learners' listening experiences. Furthermore, although VR enhanced listening experiences in both cognitive and affective terms, only the affective benefits (e.g., intrinsic motivation, sense of well-being) had a noticeable positive medium-sized effect on L2 listening comprehension. Cognitive benefits of VR, such as sustained concentration and heightened absorption, did not translate into better L2 listening comprehension. The observed relationships can be explained by the misalignment between VR's interactive elements and the cognitive demands of the listening task.

These findings have important pedagogical and practical implications. First, the study empirically supports VR's affective value in facilitating learning. Therefore, teachers and instructional designers can strategically incorporate VR into the learning curricula, especially for tasks where affective engagement is beneficial, such as introductory, motivation-building, or high-anxiety learning activities. In addition, VR learning sessions can be carefully segmented to mitigate cognitive overload without undermining engagement and motivation. Second, VR developers should design experiences in which VR's interactive elements are inherently integrated with task performance, ensuring that contextual richness enhances rather than distracts from cognitive engagement with the learning task. Third, the potential distractions of VR underscore the importance of developing learners' self-regulatory skills (e.g., effort management, attention control, and reflective thinking), so that they are less likely to be distracted by rich immersive features or tempted by hedonic activities that lead them to prioritize enjoyment over deep learning.

Despite these important implications, this study has several limitations, which future research should seek to address. First, the investigation was conducted as a single-session experiment, meaning that the effects observed may be short-lived and may not reflect outcomes over an extended period. To gain a deeper understanding of long-term impacts, future studies could incorporate repeated sessions to monitor possible changes and cumulative effects over time. Second, the VR intervention took place in a controlled laboratory environment to minimize extraneous variables, which may restrict the generalizability of our findings to real-world language classrooms. Future investigations could explore the use of VR-supported listening activities in authentic educational contexts to enhance the ecological validity of the findings. Third, listening comprehension was assessed in this study using a formal test, which may have introduced test-related effects such as test anxiety and conditioned responses. Further studies could integrate interactive assessments within VR environments to reduce test awareness and enhance ecological validity, thereby enabling a more rigorous examination of key variables. Finally, as the present findings are constrained by the scope of the study, further research could investigate whether VR's effects on listening vary across different text genres, language pairs with differing degrees of typological distance (e.g., Chinese–English vs. English–Spanish), and learner populations (e.g., L2 vs. L1 listeners). Researchers could also conduct cross-cultural replications and extend VR applications to other language skills (e.g., reading and speaking).

CRediT authorship contribution statement

Yanting Liang: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Guangwei Hu:** Writing – review & editing, Writing – original

draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Compliance with ethical standards

This study was approved by the Institutional Review Board of The Hong Kong Polytechnic University (approval no.: HSEARS20240126007)

Funding

None.

Appendices.

Appendices to this article can be found online at https://doi.org/10.1016/j.compedu.2026.105593.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Abdolrezapour, P., & Ghanbari, N. (2021). The effect of positive psychology intervention on EFL learners' listening comprehension. *Journal of Psycholinguistic Research, 50*, 1159–1180. https://doi.org/10.1007/s10936-021-09780-5

Bian, Y., Yang, C., Zhou, C., Liu, J., Gai, W., Meng, X., Tian, F., & Shen, C. (2018). Exploring the weak association between flow experience and performance in virtual environments. In *Proceedings of the 2018 CHI conference on human factors in computing systems (paper 401)*. Association for Computer Machinery.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cambridge University Press & Assessment. (2025a). B2 first for schools. https://www.cambridgeenglish.org/exams-and-tests/first-for-schools/.

Cambridge University Press & Assessment. (2025b). Quality and accountability. https://www.cambridgeenglish.org/english-research-group/quality-and-accountability/.

Chirico, A., Serino, S., Cipresso, P., Gaggioli, A., & Riva, G. (2015). When music "flows". State and trait in musical performance, composition and listening: A systematic review. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00906

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research & Development, 42*(2), 21–29. https://doi.org/10.1007/BF02299088

Council of Europe. (2020). Common European framework of reference for languages: Learning, teaching, assessment – Companion volume. https://www.coe.int/en/web/common-european-framework-reference-languages.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. Harper Perennial.

Csikszentmihalyi, M. (1993). *The evolving self: A psychology for the third millennium*. New York: HarperCollins.

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-bass.

Dahl, T. I., & Ludvigsen, S. (2014). How I see what you're saying: The role of gestures in native and foreign language listening comprehension. *The Modern Language Journal, 98*, 813–833. https://doi.org/10.1111/modl.12124

Dewaele, J.-M., & Alfawzan, M. (2018). Does the effect of enjoyment outweigh that of anxiety in foreign language performance? *Studies in Second Language Learning and Teaching, 8*(1), 21–45. https://doi.org/10.14746/ssllt.2018.8.1.2

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends, 66*, 810–824. https://doi.org/10.1007/s11528-022-00717-w

Dietrich, A. (2004). Neurocognitive mechanisms underlying the experience of flow. *Consciousness and Cognition, 13*, 746–761. https://doi.org/10.1016/j.concog.2004.07.002

Doğan, E., Şahin, F., Şahin, Y. L., Kobak, K., & Okur, M. R. (2024). Enhancing clinical law education through immersive virtual reality: A flow experience perspective. *Learning and Instruction, 94*. https://doi.org/10.1016/j.learninstruc.2024.101989

Ebadijalal, M., & Yousofi, N. (2024). "Take me to a virtual trip if you want me to write better!" The impact of Google Expeditions on EFL learners' writing motivation and performance. *Computer Assisted Language Learning, 37*, 1806–1828. https://doi.org/10.1080/09588221.2022.2123001

Egbert, J. (2003). A study of flow theory in the foreign language classroom. *The Modern Language Journal, 87*, 499–518. https://doi.org/10.1111/1540-4781.00204

Ellis, G. D., Voelkl, J. E., & Morris, C. (1994). Measurement and analysis issues with explanation of variance in daily experience using the flow model. *Journal of Leisure Research, 26*, 337–356. https://doi.org/10.1080/00222216.1994.11969966

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Green, J. C. (2007). *Mixed methods in social inquiry*. Jossey-Bass.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.

Heutte, J., Fenouillet, F., Martin-Krumm, C., Gute, G., Raes, A., Gute, D., Bachelet, R., & Csikszentmihalyi, M. (2021). Optimal experience in adult learning: Conception and validation of the flow in education scale (EduFlow-2). *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.828027

Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The Flow State Scale-2 and Dispositional Flow Scale-2. *Journal of Sport & Exercise Psychology, 24*, 133–150. https://doi.org/10.1123/jsep.24.2.133

Jackson, S. A., Kimiecik, J. C., Ford, S., & Marsh, H. W. (1998). Psychological correlates of flow in sport. *Journal of Sport & Exercise Psychology, 20*, 358–378. https://doi.org/10.1123/jsep.20.4.358

Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport & Exercise Psychology, 18*, 17–35. https://doi.org/10.1123/jsep.18.1.17

Karimi, M. N., & Nasouri, A. (2024). EFL learners' flow experience and incidental vocabulary learning during text-based game tasks: The moderating role of working memory capacity. *System, 124*. https://doi.org/10.1016/j.system.2024.103398

Keller, C., Walker, G., Amenduni, F., Tela, A., & Cattaneo, A. (2025). Find the apartment's flaws! the impact of virtual reality on vocational students' performance in general education classes and the roles of flow experience, motivation, and sense of presence. *Education and Information Technologies, 30*, 12709–12734. https://doi.org/10.1007/s10639-025-13320-2

Kim, J.-H. (2000). Foreign language listening anxiety: A study of Korean students learning English. *Digital Dissertation Consortium* (Publication No. 3004305) [Doctoral dissertation, University of Texas at Austin].

Lee, A. (2019). Using virtual reality to test academic listening proficiency. *Korean Journal of English Language and Linguistics, 19*, 688–712. https://doi.org/10.15738/kjell.19.4.201912.688

Li, C., Dewaele, J.-M., & Jiang, G. (2019). The complex relationship between classroom emotions and EFL achievement in China. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2018-0043

Li, R., Meng, Z., Tian, M., Zhang, Z., & Xiao, W. (2021). Modelling Chinese EFL learners' flow experiences in digital game-based vocabulary learning: The roles of learner and contextual factors. *Computer Assisted Language Learning, 34*, 483–505. https://doi.org/10.1080/09588221.2019.1619585

Liu, H., & Song, X. (2021). Exploring "flow" in young Chinese EFL learners' online English learning activities. *System, 96*, Article 102425. https://doi.org/10.1016/j.system.2020.102425

Liu, P., Zhang, Y., & Liu, D. (2022). Flow experience in foreign language writing: Its effect on students' writing process and writing performance. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.952044

Mahmoodi, M. H., Karbakhsh, R., & Sheykholmoluki, H. (2024). EFL learners' goal orientation, willingness to communicate, listening anxiety, and listening comprehension: A path analysis. *Journal of Language Horizons, 8*(2), 71–98. https://doi.org/10.22051/lghor.2023.43303.1796

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. MIT Press.

Mayer, R. E. (2022). Cognitive theory of multimedia learning. In R. E. Mayer, & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press.

Mayer, R. E., Makransky, G., & Parong, J. (2023). The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human-Computer Interaction, 39*, 2229–2238. https://doi.org/10.1080/10447318.2022.2108563

Moreno, R. (2005). Instructional technology: Promise and pitfalls. In L. PytlikZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-based education: Bringing researchers and practitioners together* (pp. 1–19). Information Age Publishing.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*, 309–326. https://doi.org/10.1007/s10648-007-9047-2

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*, 785–797. https://doi.org/10.1037/edu0000241

Pekrun, R. (2014). *Emotions and learning*. UN International Bureau of Education.

Pinto, D., Peixoto, B., Krassmann, A., Melo, M., Cabral, L., & Bessa, M. (2019). Virtual reality in education: Learning a foreign language. In Á. Rocha, H. Adeli, L. Reis, & S. Costanzo (Eds.), *New knowledge in information systems and technologies* (Vol. 3, pp. 589–597). Springer. https://doi.org/10.1007/978-3-030-16187-3_57.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second-language listening comprehension. *Language Learning, 55*, 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Tai, T.-Y., & Chen, H. H.-J. (2021). The impact of immersive virtual reality on EFL learners' listening comprehension. *Journal of Educational Computing Research, 59*, 1272–1293. https://doi.org/10.1177/0735633121994291

Tai, K.-H., & Hong, J.-C. (2025). Applying the Gollin effect to design VR for stargazing, and exploring the correlates between participants' VR self-efficacy, interest, frustration, flow experience and learning outcomes. *Education and Information Technologies, 30*, 1777–1799. https://doi.org/10.1007/s10639-024-12860-3

Tai, K.-H., Hong, J.-C., Chen, K.-F., & Lin, C.-L. (2024). Practicing drum on VR to promote rhythm performance: Exploring the learning progress related to incremental belief of rhythm, gameplay anxiety, flow experience, and perceived learning value. *Entertainment Computing, 48*. https://doi.org/10.1016/j.entcom.2023.100607

Tai, K.-H., Hong, J.-C., Tsai, C.-R., Lin, C.-Z., & Hung, Y.-H. (2022). Virtual reality for car-detailing skill development: Learning outcomes of procedural accuracy and performance quality predicted by VR self-efficacy, VR using anxiety, VR learning interest and flow experience. *Computers and Education, 182*. https://doi.org/10.1016/j.compedu.2022.104458

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Vertigo Games, & Knucklehead, Studios.. *Anne frank House VR* [Virtual reality application]. Anne Frank Stichting. https://www.meta.com/en-gb/experiences/anne-frank-house-vr/1958100334295482/.

Wang, X., & Feng, L. (2025). Examining the influential mechanism of English as a Foreign Language (EFL) learners' flow experiences in digital game-based vocabulary learning: Shedding new light on a priori proposed model. *Education Sciences, 15*(2). https://doi.org/10.3390/educsci15020125

Wang, L., & MacIntyre, P. D. (2021). Second language listening comprehension: The role of anxiety and enjoyment in listening metacognitive awareness. *Studies in Second Language Learning and Teaching, 11*, 491–515. https://doi.org/10.14746/ssllt.2021.11.4.2

Wu, Y.-H. S., Hung, S.-T. A., Sally Wu, Y.-H., & Alan Hung, S.-T. (2022). The effects of virtual reality infused instruction on elementary school students' English-speaking performance, willingness to communicate, and learning autonomy. *Journal of Educational Computing Research, 60*, 1558–1587. https://doi.org/10.1177/07356331211068207

Yang, B. (2021). Predicting EFL learners' achievement from their two faces—Fle and FLCA. *Theory and Practice in Language Studies, 11*, 275–285. https://doi.org/10.17507/tpls.1103.07

Yoon, H. (2025). Comparative effectiveness of small group practice and 3D VR simulation on nursing students' PIVC skills, flow state, and learning satisfaction: A quasi-experimental study. *Nurse Education Today, 150*. https://doi.org/10.1016/j.nedt.2025.106683

Zhang, X. (2013). Foreign language listening anxiety and listening performance: Conceptualizations and causal relationships. *System, 41*, 164–177. https://doi.org/10.1016/j.system.2013.01.004

Yanting Liang is a lecturer at Beijing Normal University, Zhuhai, China. Her main research interests include VR-assisted language learning and the application of AI in language education.

Guangwei Hu, PhD, is Chair Professor of Applied Linguistics in the Department of English and Communication, The Hong Kong Polytechnic University. His research interests include academic discourse, English for academic purposes, and second language education. He has published extensively on these and other areas in refereed journals and edited volumes. He is Co-Editor-in-Chief of *Journal of English for Academic Purposes*.