International Journal of TESOL Studies (2026)
260214, 1-24 https://doi.org/10.58304/ijts.260214

Article

Learner Utilization of Multimodal Feedback During Public Speaking Practice in Immersive Virtual Reality

Naiyi Xie Fincham*
University of Hawai'i at Mānoa, Honolulu, USA

Leeseul Park
University of Georgia, Athens, USA

Received: 16 Octobe 2025/Received in revised form: 29 December 2025/Accepted: 2 February 2026/Available online: 14 February 2026

Abstract
Immersive virtual reality (IVR) has the potential to foster L2 speaking development by simulating communication, enabling embodied practice, and delivering multimodal feedback. This study investigates how a commercial IVR application supports public speaking practice in the target language when app-generated feedback is paired with instructor facilitation. Intermediate-high Korean learners in a pre-capstone course used IVR to rehearse speeches for a public speaking contest. The app offered real-time feedback and performance analytics, while the instructor provided guidance informed by these analytics and their own observation. Drawing on pre/post surveys, IVR session recordings, and exit interviews, we analyze changes in learners' confidence, skill growth, and engagement with multimodal and multisource feedback. Framed by Activity Theory, our findings illustrate how VR simulations, when combined with tailored personalized feedback, cultivate holistic public speaking skills and surface implementation challenges. The study provides empirical insights into how a commercial IVR application can be pedagogically integrated in language education and offers design implications for building engaging, data-informed learning environments that provide individualized support for L2 speaking development.

Keywords
Multimodal feedback, public speaking skills, virtual reality, learner analytics, activity theory

1 Introduction

Oral presentation skills are a crucial core competence across professions and a foundational goal for higher education, as they are vital for effective communication, knowledge dissemination, and career advancement. Yet developing these skills is challenging, which involves mastering both verbal and nonverbal communication skills, overcoming public speaking anxiety, structuring and adapting

*Corresponding author. Email: naiyixf@hawaii.edu

the presentation effectively to diverse audiences and contexts (Ochoa & Zhao, 2024). For second/foreign language (L2) learners who want to pursue a professional career in the target language they are learning, developing oral presentation skills can be even more daunting, as they must develop their language proficiency and presentation skills in the target language simultaneously. Within L2 education, presentational competence is an integral part of proficiency development (Swender et al., 2012). Despite pedagogical advances, persistent constraints remain–especially in providing authentic practice environments and personalized, timely feedback. In recent years, virtual reality (VR) technology, particularly immersive VR (IVR), has emerged as a promising alternative for addressing these challenges.

VR is recognized for a range of affordances beneficial to language learning, including enabling a heightened sense of presence, embodied learning, and on-demand access to simulated scenarios and contexts that are otherwise inaccessible in physical reality (Lee et al., 2024). Additionally, specially designed VR applications are widely used in professional training in various domains to target specific knowledge areas and skills with pre-programmed learning processes and feedback, including medicine, engineering, and public speaking skills (Xie et al., 2021). For oral presentation training, not only users can practice in lifelike environments with a virtual audience but also receive holistic evaluations of their speech performance. Such applications help learners develop awareness of key performance parameters across both verbal and nonverbal dimensions through feedback delivered in multiple forms and at different stages of the practice (Van Ginkel et al., 2020). Recent advancement in IVR technology, particularly with the integration of more sophisticated artificial intelligence, expands language support, enriches learning analytics, and enables greater customization to learner needs. However, as with any educational technology, IVR's impact depends on careful design of the learning environment, instructional materials, and support for both learners and instructors (Chun et al., 2022). Particularly, given the idiosyncratic nature of L2 development and learners' interaction patterns with the learning environment, understanding how learners interpret and use feedback over time is essential to designing effective practice.

This study presents the design and implementation of L2 speaking practice with a commercial VR application in an advanced Korean language course and examines how the application can support the development of public speaking skills in the target language. The study specifically focuses on the impact of personalized facilitation, consisting of app-generated multimodal feedback and instructor guidance. Although the target language here is Korean, the lessons learned from the design, integration, and implementation of this technology, as well as the findings on students' use of multimodal and multisource feedback have implications that can be extended to instructional settings for other languages.

In what follows, we review related literature that informs our practice design and research, describe the instructional context and learner experience, detail the research setting and methods, and then present findings before turning to implications and conclusions.

2 Literature Review

In this section, we review literature in areas relevant to the present study: (1) developing oral presentation skills in an L2, (2) IVR and multimodal learning analytics for public speaking development, and (3) a sociocultural account of language learning operationalized through Activity Theory, which provides the theoretical framework for our research design.

2.1 Developing oral presentation skills in an L2

Oral presentation skills are considered essential for an L2 learner's proficiency development, as reflected in all established proficiency guidelines (e.g., American Council on the Teaching of Foreign

Languages [ACTFL], 2012; Council of Europe, 2001). Thanks to decades of L2 research in speaking development and pedagogical insights, we are well informed about what types of tasks, feedback and practices are effective for developing oral presentation skills. First of all, presentation tasks need to be designed to mirror real-world use and require students to integrate all four language skills in a natural way, for example, in-class collaborative or individual project presentations, and simulated scenarios such as debates, business pitches, and speaking competitions (e.g., Wang et al., 2018; El Majidi et al., 2021; Beltrán-Palanques, 2024; Le Vo, 2022). Implied in these real-life-like tasks is the need to create an audience presence. Research has shown that practices and rehearsals in front of an audience are positively correlated with presentation quality, as the audience's reaction can serve as important nonverbal feedback to the speaker (Carrell & Menzel, 1997). Another key task design principle is to make the evaluation criteria salient to the learners, which is instrumental in helping learners to set goals and to self-assess their presentation performance (Weimer, 2013; Khonamri et al., 2021).

Feedback is another essential component for a successful oral presentation pedagogy. Research has examined the effects of feedback in different aspects, including types of feedback, the medium or technology used to deliver feedback, and the timing of feedback. Specifically, multimodal feedback is essential in enhancing speaking skills through holistic speech training, which incorporates various modes of communication (linguistic, visual, auditory, kinesthetic) to provide comprehensive feedback to different dimensions of oral presentations (Guichon & Cohen, 2016). This is supported by the embodied cognition perspective that cognition and learning are shaped by bodily states and situated action, rather than being purely "in the head" (Wilson, 2002; Barsalou, 2008). Applied to L2 development, this view foregrounds language use as coordinated perception-action in context (e.g., posture, gesture, gaze, breathing and prosody), making speaking an embodied experience (e.g., Atkinson, 2010). Therefore, multimodality is considered particularly beneficial for learners to develop a comprehensive understanding of the target language in action, especially how it is used for different communicative purposes or in different sociocultural contexts (Perez, 2020; Jiang & Hafner, 2024).

In terms of feedback sources, overall, research in speaking skill development indicates that feedback from the teacher remains the most effective, outperforming peer and self-feedback in improving presentation performance (Garbati & Mady, 2015). However, orchestration of feedback sources has shown to also have strong effects on performance gains. For example, a 2024 study (Liu & Aryadoust, 2024) on public speaking classes reported that the group of students who received the combination of peer and teacher feedback showed the largest gains in their performance as well as behavioral engagement.

Feedback-giving technology, i.e., how feedback is provided to learner can also affect the quality of the feedback-learning process. Research shows that feedback delivered through video-based reflection is an effective method that can enhance public speaking pedagogy. Studies have found that providing students, L2 learners in particular, with their recorded performance with additional commentary can increase reflective depth, self-awareness of strengths and weaknesses, and self-regulation (e.g., Tschirner, 2001; Miskam & Saidalvi, 2020). Additionally, being able to observe themselves on video has also shown to improve voice-related aspects of speech delivery, especially pronunciation (Sun & Yang, 2013; Tatzl, 2017).

In terms of timing of feedback, i.e., immediate (real-time) feedback vs. delayed (post-task) feedback, research has shown advantages and disadvantages in both. For instance, immediate feedback during speech may lead to in-the-moment behavioral adjustment, while delayed feedback from experts may facilitate long term retention of skills (King, Young & Behnke, 2000). These findings imply that it would be ideal to integrate both feedback mechanisms in the actual practice experience.

In summary, research indicates that real life-like scenarios, teacher feedback, and multimodal feedback including real-time objective analytics are key factors in developing public speaking skills. Yet they place substantial human and logistical demands on instructional contexts, particularly when

coordinating feedback forms and timing (Van Ginkel et al., 2015; Van Ginkel et al., 2020). To address these challenges, researchers and educators have leveraged technological advancements, including VR-based tasks, as potentially viable alternatives or supplements for supporting practice and feedback.

2.2 IVR and multimodal feedback for L2 public speaking practice

IVR offers a computer-generated, spatially realistic environment experienced through a head-mounted display and hand-held controllers, creating a strong sense of presence and embodied interaction (Dhimolea et al., 2022). In language education, IVR has been discussed as a promising medium because it can increase engagement and support situated practice by combining immersion, multimodality, and interactivity (Chun et al., 2022; Lee et al., 2024; Kaplan-Rakowski, 2025). For the purposes of oral communication, these affordances are particularly relevant because learners can rehearse performance in contextualized settings while managing affective factors such as anxiety and self-consciousness (Thrasher, 2022; Kaplan-Rakowski & Gruber, 2023).

This is especially pertinent to L2 public speaking and oral presentation pedagogy, where skill development depends on repeated practice with authentic tasks, the psychological presence of an audience, and access to actionable feedback—conditions that are often difficult to provide consistently in classrooms due to scheduling and resource constraints (Van Ginkel et al., 2020). Public-speaking–oriented IVR applications can simulate realistic venues (e.g., stages, lecture halls) and variable audience sizes, allowing learners to rehearse in settings that resemble their target performance context or to gradually scale difficulty as they build confidence (Wörtwein et al., 2015). Virtual audiences can also elicit performance-relevant psychological responses similar to those triggered by real audiences, thereby making rehearsal more consequential than practice without an audience (Seuling et al., 2024).

A distinctive contribution of IVR public speaking platforms is their capacity to provide multimodal feedback and analytics that capture the multidimensional nature of speaking performance. Systems can integrate visual, auditory, and haptic cues to target verbal and acoustic features (e.g., speech rate, pauses, filler words) alongside nonverbal behaviors (e.g., gaze/eye contact, gesture, body movement). For instance, learners may receive on-screen prompts to adjust pacing or eye contact, controller vibrations signaling hand movement, or post-task dashboards summarizing delivery metrics. Such multimodal feedback can support holistic speaking development by making otherwise tacit aspects of performance perceptible and actionable for self-regulation (Ochoa & Zhao, 2024). Recent AI-enabled advances have further increased the granularity and responsiveness of these analytics, enabling more individualized performance feedback in commercial tools (e.g., Ovation VR, VirtualSpeech).

Despite these potentials, we still lack detailed accounts of how learners attend to, interpret, and take up different feedback types across repeated IVR practice, and how feedback engagement relates to subsequent performance. Addressing this gap, the present study examines learners' public speaking practice processes in an IVR environment, focusing on how multimodal feedback is utilized during and after practice sessions and how it shapes preparation for the final speech.

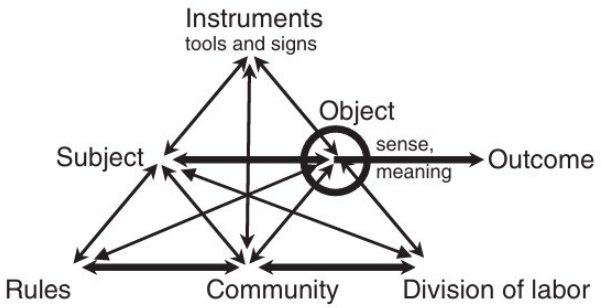
2.3 Activity theory as analytical framework

In this study, we adopt a sociocultural perspective grounded in Vygotskian theory (Vygotsky, 1978), which holds that cognitive development is shaped by the social and cultural-historical contexts in which learners participate. From this view, thinking, learning, and problem solving are mediated through social interaction and cultural tools – both symbolic (e.g., language and signs) and material (e.g., technologies and tools). Human development thus emerges through mediated activities coordinated among learners, other collaborators, and artifacts and their culture of use (Thorne, 2003). To operationalize this perspective, we use Activity Theory to analyze how different elements within an activity or activity system coordinate and evolve.

Activity Theory (AT), particularly the third-generation AT and Engeström's account of expansive learning (Engeström, 2014), takes the collective, objective-oriented activity system as the unit of analysis and examines the historically developing relations among subjects, mediating artifacts, community, rules, division of labor, objects (goals), and outcomes (Engeström, 1987, 2014). This lens enables us to analyze how tools, roles, and contextual conditions jointly shape action, learning, and system change.

Engeström's (2014) general activity system model (Figure 1) depicts the dynamics and interconnectedness of the components (nodes) in an activity system: the subject(s) engaged in the activity; the object(s) or goal(s) that drives it; the instruments (tools and signs) that mediate action; rules that regulate participation and communication; the community in which the activity occurs; and the division of labor that distributes roles and responsibilities; and together they result in the outcome(s) (Engeström, 2014; Nocchi, 2018).

Figure 1
General Model of an Activity System (Engeström's, 1987)



Drawing on research grounded in AT, our study centers on two core mechanisms: mediation and contradictions. Mediation refers to how humans use instruments (tools, artifacts, signs) to interact with their environment in order to make sense of the world and achieve their goals (Lantolf, 2006). In IVR-supported language learning, mediational means include material tools such as the VR headset and controllers, the monitor used for screencasting, and in-app resources (e.g., slides, timers, pointers, avatars), as well as symbolic tools such as the instructional materials and target languages, task instruction and rubrics, and the multimodal literacies the activity demands. Social mediation is provided by instructors, peers, and other participants through modeling, scaffolding, and feedback.

As activity systems are affected by the context they function and are constantly changing and evolving, the mediational forces among elements within the system and the interaction between connected activity systems can give rise to contradictions or tensions (Nocchi, 2018). Contradictions are "*historically accumulating structural tensions within and between activity systems"* (Engeström, 2001, p.7) that arise as the system develops. They can occur within a single node of the activity system (primary), between nodes (secondary), between the current form of the activity and a more advanced new form (tertiary), or between the focal system and neighboring systems (quaternary) (Engeström, 2014). In Engeström's view, contradictions are the "necessary but insufficient" drivers of change within activity systems. They prompt participants to question and reconceptualize the object, opening paths to collective re-mediation and expansive learning. Tracing how contradictions emerge and are addressed can reveal leveraging points for redesign (e.g., tool configuration, rules, division of labor) and for refining facilitation and analytics.

In this study, we analyzed students' "response actions" within the activity of practicing public speaking in IVR. We analyzed how the different components within this activity system (i.e. subject, object, tools, rules, community, and division of later) would mediate students' speech rehearsal and utilization of the multimodal feedback as they completed a series of guided practice sessions in IVR. We also identified contradictions that have surfaced when practice broke down or deviated from expectations,

revealing tensions and barriers among system elements. Specifically, the present study was guided by the following research questions (RQ):

RQ1: How did students' feedback-response behaviors evolve across IVR practice sessions when receiving multimodal feedback?

RQ2: How were the availability and students' utilization of different feedback types associated with their final speech performance?

RQ3: What contradictions within the IVR activity system emerged around feedback utilization and public speaking practice, and how did students navigate these contradictions?

The next section describes the instructional context and design of the IVR practice session, outlining two interrelated and interacting activity systems. This is followed by the research design, presentation of findings, discussion, and implications.

3 Instructional Context

In this study, the instructional context is the Korean Language Flagship Program, a U.S. federally funded initiative that cultivates professional-level proficiency in Korean and prepares learners for effective participation in global academic and professional settings. The focal instructional site was an advanced Korean course that emphasized oral proficiency in professional communication. Specifically, developing public speaking skills in Korean was a key objective of this course. Like L2 learners of other target languages (TL) who have very limited contact with the TL community outside the classroom (Ding, 2024; Yoon & Seo, 2024), students in Korean programs frequently face parallel challenges: limited opportunities for extended oral practice, lack of explicit instruction in public speaking, minimal experience performing before an audience, and heightened speech anxiety. These challenges underscore the pedagogical importance of designing learning experiences that provide sufficient rehearsal opportunities, constructive feedback in supportive, low-stakes practice environments. Addressing these needs is essential for enabling learners to transition from classroom-based communicators to confident, professional-level speakers capable of communicating with diverse audiences.

The objective for this course was to prepare students for their one-year capstone experience in Korea, which included a six-month internship at a Korean company or organization. With this goal, the 16-week course was created around a simulated experience originally designed by the Language Flagship Technology Innovation Center (Tech Center) (Fincham et al., 2022) where students applied for jobs at a fictitious multinational consulting company and once hired, completed required training for new employees, all in the target language. The curriculum was organized following the Tech Center's design, which includes two modules: (1) a Job Application module, which guided learners in preparing application materials and conducting mock interviews; and (2) an Onboarding module, which addressed cultural expectations and norms in professional settings, and public speaking skills, which is a highly regarded skill in Korean business culture. Both modules shared the broader instructional goal of aligning linguistic proficiency with practical, real-world communicative demands in professional contexts. The final project for the course was a public speaking competition modeled on the Ignite® Talk format, in which each student delivered a five-minute speech accompanied by 20 auto-advancing slides. This format requires learners to integrate accuracy, fluency, timing, and rhetorical strategy under real-time communicative pressure in front of a live audience.

The decision to use IVR in this study was based on instructor's observations and needs analysis conducted in previous years' implementations. Students shared that they wanted more practice opportunities, a low-anxiety and low-stakes environment for practice, and personalized feedback on their performance. To meet these needs and considering the available resources and constraints within the institution, a single-user IVR application (Ovation VR) was selected. This application offers

functionalities that are aligned with the course goals and learners' needs. It provides repeated exposure to simulated public speaking scenarios, enabling learners to confront their speaking anxieties while rehearsing in less stressful environments. The system supports 21 languages including Korean, and processes language-specific linguistic features such as filler words. Users can select multiple practice venues, such as classrooms, conference halls, and auditoriums. Audience characteristics are configurable as well, including audience size and overall attitudes (positive, negative, or neutral). Speaker aids, such as teleprompter and timer, are also available to help students when they are still working on memorizing the speech or when they need to practice time management. With this application, students can rehearse in environments that were initially less intimidating and progressively more challenging, thereby learning to manage speech anxiety. Most importantly, the application generates feedback through real-time performance tracking of speech rate, gaze, and usage of filler words and hand gestures, offering analytical summary, allowing students to integrate feedback across multiple practices.

4 Research Design

A qualitative case study design (Creswell & Poth, 2016) was adopted for this study because our aim was to develop an in-depth, contextualized understanding of how individual learners engaged with and acted on different feedback types across repeated IVR practice sessions. This design allowed us to trace within-case processes over time and relate feedback uptake to the local activity system shaping learners' practice.

4.1 Participants

The current study received institutional ethics approval through the Institutional Review Boards prior to data collection. All participants provided consent via consent form and were informed that their participation was voluntary and would not affect their course evaluation. The project involved five undergraduate students who were enrolled in the Korean Language Flagship Program. Their Korean proficiency ranged from Intermediate to Advanced, as determined by placement within the Korean Flagship curriculum and instructional benchmarks aligned with the ACTFL scale (ACTFL, 2012). Out of the five students, two students are selected as focal participants for the present study, based on the completeness of their data across all IVR practice sessions and their final speech performance. Student 1 (S1) is a non-heritage learner pursuing a double major in Health Science and Korean. At the time of the study, this student had three years of Korean learning experience, all with the Flagship program. Student 3 (S3) is also a non-heritage learner, majoring in Korean with the Flagship program. This student had a more extensive background in Korean, with a total of six years of Korean language learning experience, including formal instruction in high school. Regarding their experience with IVR, the pre-survey responses indicated limited prior exposure. Both students hoped the practice would provide immediate, structured feedback in a controlled environment, increase their comfort with presenting to an audience, and offer a realistic rehearsal opportunity to refine timing, pacing, and speech length. As for the course instructor, although she served as a researcher in the study, several safeguards were implemented to manage the instructor-researcher dual role. These included the practice protocols, consistent feedback procedures applied to all students, and conducting analytic coding and interpretation after course completion.

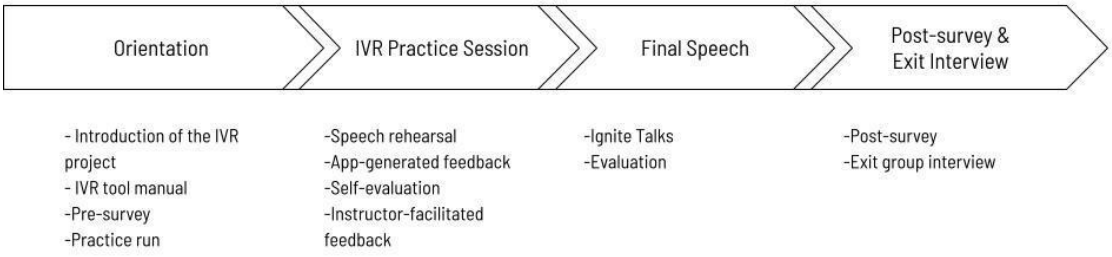
4.2 IVR practice session

The pre-capstone course met once a week for 75 minutes over 16 weeks. Students began working on their final speech once they had started with the Onboarding module. Their final speech performance was to be evaluated by two judges, one native and one non-native speaker of Korean, both with extensive

experience teaching Korean at the university level in the United States. The analytic rubric (Appendix B) was developed to reflect the course objectives and student learning outcomes related to public speaking skills.

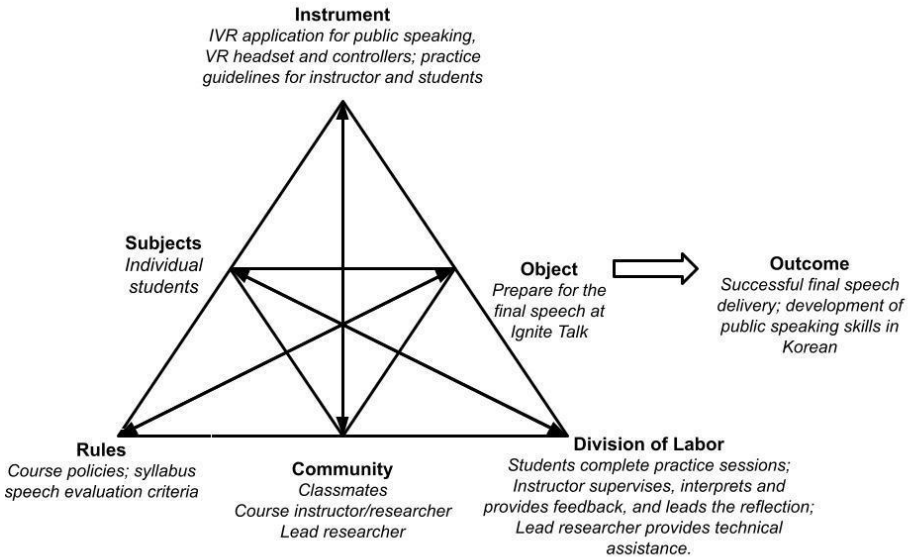
IVR practice sessions were conducted using a single-user immersive VR setup with the Meta Quest 2 headset and Ovation VR. Sessions took place in a studio equipped for audio and video recording, allowing sufficient space for head and hand movement while wearing the headset. To maximize practice effects (Boetje & van Ginkel, 2021), each student was encouraged to complete at least three IVR practice sessions. Each session consisted of a speech rehearsal in a simulated venue followed by review of app-generated feedback and instructor-led reflection. These technical and procedural configurations shaped the affordance of the IVR environment, including the simulated audience and the availability of real-time feedback. IVR practice began in the second half of the semester after the Job Application module, facilitated by the lead researcher and the second researcher (the course instructor). A preliminary needs analysis informed the design of instructional materials, guidelines, feedback procedures, and pre- and post-surveys. To ensure consistency, the researchers created a practice session protocol that included detailed step-by-step notes on procedures, feedback provision, reflection activities, and monitoring. The complete data collection is organized in four phases (see Figure 2 for the flow of activities involved):

Figure 2
Flow of Activities Involved in the IVR Practice Experience



Guided by the AT analytical framework, the focal activity system is thus the IVR practice session, consisting of the following components as depicted in their initial status in Figure 3:

Figure 3
Elements in the Focal IVR Practice Activity System



4.3 Data collection and analysis

To address the RQs with the AT framework, multiple sources of data were gathered, including screencast recordings of each IVR practice session, pre- and post-evaluation surveys, instructor feedback logs and notes, focus group interview transcript, judge's evaluation of final speech delivery, and all related instructional and task materials. Specifically, the types and sources of multimodal feedback are summarized in Table 1.

Table 1
Multimodal Feedback during IVR Practice Session

Modality/Source	From IVR App		From Instructor	
	During	After	During	After
Text + Visuals Integrated	Gaze Speech rate Voice Hand movement Filler words Virtual audience reaction Speech rate Voice	Summary analytics in numbers, percentages, pie charts and graphs	x	x
Audio/Verbal	x	x	Reminder and navigation assistance	Pronunciation Word choice

Analysis of the IVR session recordings was carried out using an iteratively developed coding scheme to capture patterns of speech performance development, feedback utilization, and changes in individual learner variables (confidence, self-assessment, and perceptions). The initial coding scheme was constructed based on research on feedback and feedback uptake in L2 context (Lyster & Saito, 2010) while considering the medicinal forces within the system. The codes are organized into four main categories (see Appendix A for the full codebook):

- Tool-Mediated Responses: learners' interaction with app-generated feedback, ranging from immediate use to misinterpretation, reflection, or rejection
- Instructor Feedback Uptake: learners' responses to instructor guidance, including immediate uptake, delayed uptake, reflection, and rejection
- Tool Discussion: learners' evaluations of VR features, coded as positive or negative reflections on usability and affordances
- Outcomes: documented changes in confidence, self-evaluations of learning, observable performance improvements, and challenges such as time management

The initial coding scheme was piloted using two IVR session recordings from a third participant. The first and second researchers independently coded this pilot dataset to test code definitions and boundaries. Percentage agreement for the pilot independent coding round was 89%. The researchers then met to review all instances of disagreement, discussed rationales for code selection, refined code definitions and decision rules, and updated the codebook accordingly. Due to logistical constraints (limited

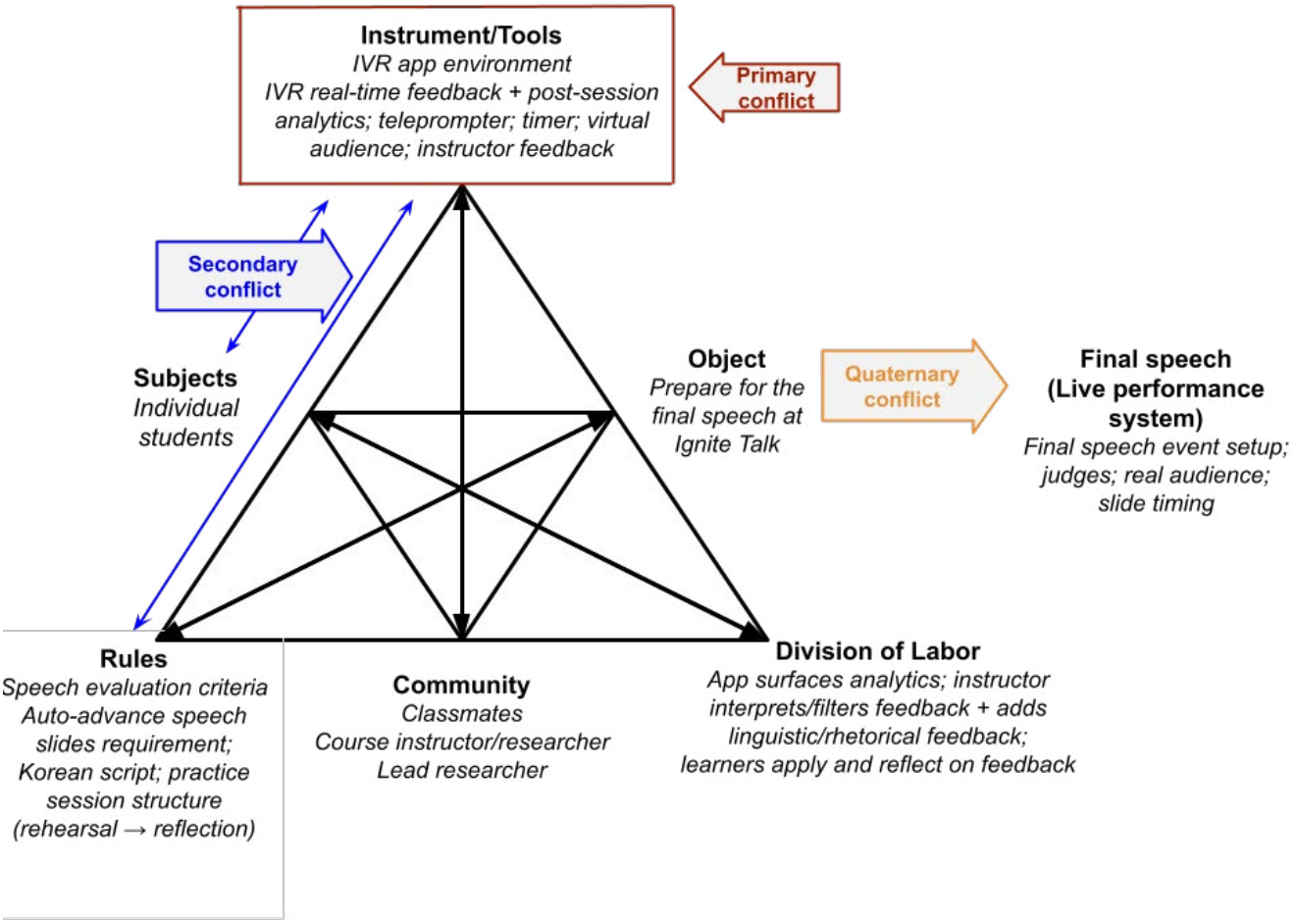
access to analytical software), the full dataset for the two focal participants was subsequently coded collaboratively by the first and second researchers using the revised codebook. Throughout this process, any uncertainties or discrepancies were documented and resolved through discussion to reach consensus, and code definitions were iteratively clarified when needed.

Responses in the pre- and post-evaluation surveys completed by the two focal students were analyzed thematically using the same coding scheme. These results, together with the instructor's field notes and related instructional materials were then organized and synthesized with the AT model to identify changes and emerging patterns.

5 Findings

Drawing on the AT analytical framework, we report findings regarding patterns of feedback use, feedback availability, and how they relate to students' final speech performance. We also report moments where practices were not going smoothly, revealing contradictions among system elements and how learners navigated through adaptive strategies. Figure 4 maps the reported findings to AT components and contradiction levels. Detailed findings are presented below according to the research questions.

Figure 4
Summary of Findings Mapped to the Focal IVR Practice Activity System



5.1 Utilization of multimodal feedback during IVR practice sessions

Findings regarding the utilization of multimodal feedback are organized in three sub-categories by feedback source: app-generated feedback, instructor feedback, and IVR in general.

5.1.1 App-generated feedback

Overall, both focal students' utilization of app-generated feedback changed over the three sessions in terms of responsiveness and frequency and varied by feedback type. During the first practice sessions, most of the app-generated feedback related to speech performance itself was not attended to by either S1 or S3. The time reminder feedback was one that received the most response. Upon seeing the time-remaining message, students would immediately speed up their speech in order to finish within the time limit. Students later all agreed that being able to practice with the timer from the very beginning helped them pace their speech and adjust the speech content. It was not until the reflective activity led by the instructor did the student become aware of the other speech related feedback during their rehearsal. As shown in Figure 5 (the scarcely distributed stripes in the orange highlighted part), participant S3 ignored all of the app-generated feedback during the speech rehearsal, but engaged in discussion of the app-generated feedback during the self-evaluation and instructor feedback period. Specifically, during self-evaluation and instructor feedback, S3 took up the app-generated feedback regarding gaze and hand movement either following instructor's explanation (yellow strips: APP_INSTR_GUIDED_POST), or self-reflection (pink strips: APP_REFLECTIVE_USE; teal strips: APP_ACCEPTED_DELAY).

Figure 5
Coding Summary of Speech Rehearsal #2 in the First Practice Session (T1) of S3

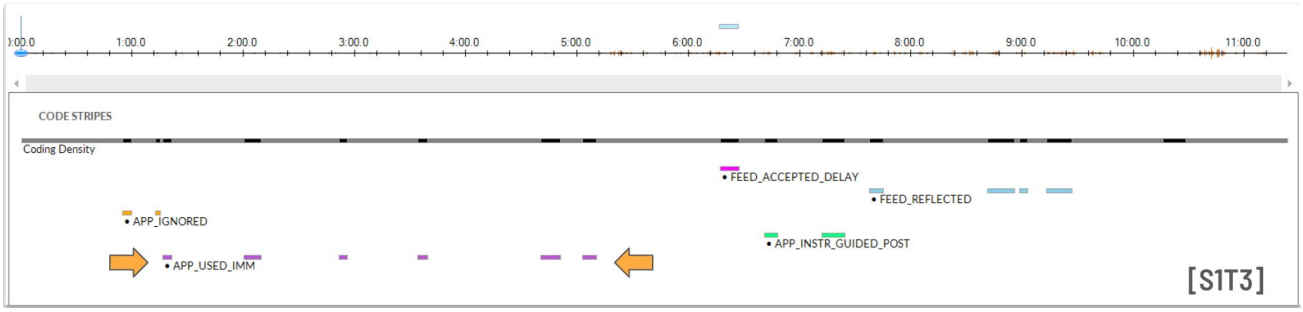
She also talked about this experience in the exit interview:

> …[I] remember the feedback category gaze and hand gesture and that was helpful at the very end, but the things that popping up while I was speaking just threw me off more honestly I'm interrupting more, so you cannot focus… [S3, Exit interview]

While this pattern continued in the following two practice sessions, there was a visible increase in students' utilization of speech-related app-generated feedback during speech delivery in later practice sessions. Starting from the second practice session, both students paid attention to the real-time feedback during their rehearsals and responded to some of them either by adjusting their speech immediately, adjusting following explicit acknowledgement (e.g., verbally like 'ok' or nonverbally like nodding) of the feedback, or ignoring the feedback after acknowledgement or verbal commenting. As shown in Figure 6, participant S1 had more immediate uptake of app-generated feedback during her rehearsal (pink strips: APP_USE_IMM), including adjusted speech rate, voice volume, and hand movement.

Figure 6

Coding Summary of Speech Rehearsal #1 in the First Practice Session (T3) of S1



In terms of the portion of app-generated feedback that was ignored or rejected during the self-evaluation and instructor feedback phase, the reason was mainly due to inaccuracy or inapplicability. One of the most rejected feedback types was related to the use of filler words. The app either detected filler words while they were not produced, or the instructor believed that the detected filler words were acceptable given the actual context, for example, when signaling a brief pause without interrupting the speech flow. See Figure 7 for a summary of students' action in response to app-generated feedback.

Figure 7

Summary of Students' Actions to App-generated Feedback

Name	Description	Files (total 11)	References
Tool-Mediated	**Learner response to app-generated feedback (real-time and summary)**	**7**	**72**
APP_ACCEPTED_DELAY	Learner applies app feedback in later sessions or final speech	3	6
APP_IGNORED	Learner does not respond to app-generated feedback	6	17
APP_INSTR_GUIDED_POST	Instructor interprets/guides students through app generated feedback during review	3	9
APP_MISUNDERSTOOD	Learner misinterprets the app feedback, applies it incorrectly	1	1
APP_REFLECTED	Learner discusses or evaluates app-generated feedback	4	4
APP_REFLECTIVE_USE	Learner reflects on the role of the tool in their learning (from review, interviews or surveys)	4	16
APP_REJECTED	Learner chooses not to act on app generated feedback, possibly with reason	3	5
APP_USED_IMM	Learner uses app-generated feedback immediately	4	14

5.1.2 Instructor feedback

Across the repeated IVR practice sessions, instructor feedback emerged as the main mediational resource that helped learners interpret and act on app-generated multimodal feedback. In doing so, the two focal students were able to redirect their attention to some overlooked app-generated feedback and integrate the feedback in next speech rehearsals. The instructor also provided individualized feedback on areas beyond the scope of IVR analytics, including target language-specific pronunciation and pragmatic approaches to word choice. Such feedback again facilitated self-awareness and behavioral adjustment. For example, when Student 1 received the instructor's feedback on pronunciation, she immediately acted on it by practicing the words multiple times to improve her pronunciation. Her uptake of the feedback was evident in her next IVR session, where her pronunciation of those words showed noticeable improvement. Figure 8 presents another example of Student 3's active engagement with instructor-facilitated feedback. After her first try during that session, the instructor provided individualized feedback, highlighted key app-generated feedback (e.g., gaze patterns, balancing hand movement), and filtered out some misidentified app-generated feedback, such as filler words in Korean. As the orange-marked segments in the figure show, participants accepted and discussed the instructor's feedback and accordingly adjusted for their next rehearsal.

Figure 8

Coding Summary of Speech Rehearsal #1 in the Second Practice Session (T2) of S3

Figure 9

Summary of Types Instructor Feedback and Students' Update

Name	Description	Files (total 11)	References
Instructor Feedback Uptake	**How students respond to feedback from instructor after IVR practice**	**6**	**28**
FEED_ACCEPTED_DELAY	Learner applies instructor feedback in later sessions or final speech	2	3
FEED_ACCEPTED_IMM	Learner takes immediate action based on instructor feedback (e.g., adjusts gestures mid-practice)	3	9
FEED_REFLECTED	Learner discusses or evaluates instructor feedback	5	14
FEED_REJECTED	Learner chooses not to act on/accept instructor feedback, possibly with reason	2	2

Figure 9 shows a consistently high level of uptake of instructor-facilitated feedback, distributed across four response types. Many cases of "FEED_REFLECTED" are concerned with areas requiring further discussion, such as word choice. For example, Student 3 received feedback on her use of loan words and expressed uncertainty about whether to replace them with Korean equivalents. This led to further reflective discussions on these words and extended to other related words or context, which sometimes resulted in the student using the alternative word choices in her next speech. Instances of FEED_ACCEPTED_IMM were associated with features that were comparatively easy to adjust, such as tone, in which cases students usually incorporated the corrected tones immediately.

5.1.3 IVR as practice environment

Throughout the practice sessions, both students commented and reflected on the different features and functionalities of the IVR environment in general and the application in particular, and how they affected their speech development and overall experience, both positively and negatively. On the positive side, students felt that the reminder of body language such as gaze, hand movement, as well as remaining time were more helpful as they could make timely adjustments while speaking. As Student 1 commented:

> "Because if I was just doing it alone in … in the room by myself, I'd see the reflection be like, Okay, I gotta move my arms more, and I've gotta not mess up. Let's try it again. But you were able to be like, oh, let's look at this section specifically. " [S1, Exit interview]

However, they found feedback related to their speech rate and filler words were more distracting than helpful during the speech, and that feedback as such 'only made sense' during the reflection, as Student 3 commented:

> … "that was helpful at the very end, but the things that popping up while I was speaking just threw me off more honestly, I'm interrupting more so you cannot focus on saying what all it's like. I know I'm stuttering. I just don't remember what the word is. It wasn't helpful in that moment, but afterwards, when I saw it and I was able to see when I was reading over what I read, I said all three times in a row." [S3, Exit interview]

5.2 Multimodal feedback and final speech performance

By examining students' feedback engagement patterns across IVR sessions, their self-report, alongside judges' evaluations, it becomes clear that students' ways of acting on feedback have contributed to their confidence, delivery, and rhetorical effectiveness during the final speech performance. At the same time, we do acknowledge that these relationships should be interpreted alongside other learning opportunities in the larger course context and additional rehearsal support they may have received (e.g., coaching prior to the live event).

Student 1's engagement with multimodal feedback shifted from ignoring app-generated feedback to increasingly higher uptake, involving both immediate adjustment and reflection. By the third IVR session, she combined immediate adjustments (e.g., rehearsing pronunciation directly after receiving instructor correction) with delayed incorporation (e.g., refining delivery features such as pacing and gesture in subsequent rehearsals). The benefits she gained from incorporating the feedback were reflected directly in her final speech, where judges praised her confident and natural delivery, natural body language, and accurate language use. She received near-perfect scores (54/55), with full scores in organization, content, language, and delivery. While she initially reported low confidence during the first session, she later described a clear increase in confidence in both the post-survey and the focus group interview. She explained that the repeated IVR sessions helped her feel more comfortable with public speaking and improved her ability to control her voice volume.

In a similar pattern, Student 3 often ignored or rejected app-generated feedback in the early IVR sessions. Her engagement increased during the reflection, when the instructor provided interpretation of the app analytics and highlighted areas for improvement, particular regarding filler words, body movement, word choices). As a result, she gradually integrated both app and instructor feedback across sessions, through active discussion with the instructor. Her final speech demonstrated clear improvement in organization, story development, and visual design. Judges further complimented her clear voice and pronunciation, as well as her effective use of visual aids (5/5), noting how she successfully expanded her topic from a personal to a broader social dimension. Content and organization scores (9–10/10) reflected these strengths, and her visuals were described as polished and engaging. In her post-survey and focus group interview, Student 3 reported that IVR practice sessions were valuable for reducing speech anxiety and providing opportunities for rehearsal. Specifically, she explained that practicing with a virtual audience helped her feel less intimidated by live audiences, a shift from her initial anxiety when facing the virtual audience in earlier sessions. In sum, her case illustrates how delayed, reflective uptake of multimodal feedback, supported by the instructor's guidance, can strengthen both linguistic (e.g., pronunciation) and non-linguistic aspects (e.g., voice, speech anxiety).

5.3 Emerging contradictions

Our analysis identified three types of contradictions, in other words, tensions or barriers that helped explain why things did not work or did not work as expected. These contradictions emerged at different levels and between different nodes within and across activity systems, including the tool itself, and between tools and rules of the task and the course project.

5.3.1 Contradictions related to tool functionalities and subjects (participants)

These primary and secondary barriers result from some of the IVR app's functionalities that affected students' practice experience negatively at the interface of personal preferences. First, while the presence of the virtual audience was there to make the simulated speech feel more authentic, it turned out to be disturbing to some students. This was the case for Student 3, who was so much aware of the lack of facial expressions and varied body movements displayed by the virtual audience that she felt that her speech was not good enough, all the while knowing that the audience themselves could not process the speech content. As she explains in the following excerpt:

> "They're just like shaking their leg or whatever, the people in the audience are, like pointing, or they're kind of whispering, or they're on their phone, or there's like a flash coming from some side or distracting." [S3, Exit interview]

On the other hand, tension that was created by having a virtual audience presence also promoted students to use adaptive strategies which were then transferred to the final speech event, for example, practicing how not to get distracted or intimated by the audience by choosing to a focal point away from the audience. The following excerpt from the exit interview reflects students' views on this:

> [10:34]: Yeah, I mean, it's AI, but I wasn't really trying to look at them anyway because my whole point was just I'm just gonna stare at the clock and I'm gonna look at nothing else but the clock.
>
> [10:43]: That's what I did in the real thing too.
>
> [10:44]: The second I made eye contact with the person is when I started fumbling, and so I realized like I've just got to not look at them to be able to succeed.
>
> [Group interaction. Exit interview]

Further, although Korean is one of the languages supported by the app, in our project, this function seemed not adequately stable to accurately detect Korean filler words, which caused some confusion and unwanted distraction. While the instructor was able to clarify with the students during the reflection, this limitation had led to some doubts among the students regarding the feedback's reliability. Consequently, feedback related to Korean filler words was rejected the most and ignored most of the time across all practice sessions.

Lastly, feedback related to hand movement was deemed less useful, as students considered it "not realistic". To students, having to hold the controller during the speech rehearsal was far less authentic than the actual hand movement they could do on stage. Therefore, while they acknowledged the importance of hand gestures during speech, they were not able to realistically practice that in the IVR setting.

5.3.2 Contradictions related to tension between tools and rules

The second type of secondary contradictions emerged between the IVR app's functionalities and design and the task rules. It was part of the course's requirements that students should write their speech scripts in Korean. However, while students were able to upload their script into the IVR app, the Korean fonts sometimes were not displayed correctly in the teleprompter, which had a negative impact on the students' speech rehearsal. They had to make extra efforts to decipher what the mis-displayed characters were while speaking, at the same time trying to attend to other aspects of the speech. This issue was brought up constantly by the students during the first two practice sessions when students still needed to rely on the script to complete the speech, and again during the focus group interview when they shared their experience with the IVR app.

Next, while the IVR app does provide the option for users to upload their presentation in PowerPoint format and use it during the presentation, the Ignite talk format requires that the presentation slides be automatically advanced every 20 seconds, which is not supported by the app. As a result, students could not upload their presentations and practice their speech in coordination with the slides. This was deemed a downside by the students, who shared comments such as the following:

> "Yeah, if it could move every 15 seconds. That'd be great, because then we can actually check. It wasn't until I actually went in for the day before the rehearsal with my speech coach, where I was able to, like, have the slides on the board behind me too, yeah, actually start talking. I was like, okay, when I say this, this is what slide comes up, yeah. And that would have been really helpful."

Clearly, the lack of such functionality created a mismatch between the simulated practice with the particular reality that the students will face, i.e., the Ignite Talk format with automatically advanced slides. The impact of such misalignment was later reflected in the judges' evaluations regarding the "inadequate 'coordination between speech and presentation visuals" during the final speech, which was likely due to the insufficient practice.

6 Discussions

RQ1: How did students' feedback-response behaviors evolve across IVR practice sessions when receiving multimodal feedback?

Our analysis shows that in earlier practice sessions, students largely ignored real-time app-generated feedback during rehearsal. This may be interpreted through the lens of cognitive load in multimedia learning. Building on cognitive load theory, Mayer's (2005) Cognitive Theory of Multimedia Learning (CTML) argues that learning with multiple modalities is constrained by limited working memory and dual processing channels (auditory/verbal and visual/pictorial). In our context, although students completed a brief orientation before Session 1, once the speech began, they still had to navigate the IVR environment and coordinate parallel processes: delivering content, tracking the teleprompter, monitoring the timer, orienting within the virtual venue, and operating features via head and hand controllers. With all these demands competing for attention, continuous on-screen prompts added to extraneous processing, thus were often deprioritized. As students became more familiar with the practice environment, extraneous load decreased, allowing more resources for germane processing of feedback; correspondingly, we observed greater uptake of real-time cues in later sessions.

Importantly, even when students did not engage with real-time prompts, they used app-generated analytics productively during instructor-led reflection. Post-speech summaries (e.g., pacing, gesture frequency, filler-words counts) served as externalized attention guides that supported self-evaluation and targeted instructor feedback. Within the Activity Theory framework, this pattern reflects an emergent division of labor: the instructor functioned as the primary source of linguistic feedback (content, language use, pronunciation), while the application provided multimodal performance analytics (nonverbal and delivery parameters).

RQ2: How were the availability and students' utilization of different feedback types associated with their final speech performance?

Across sessions, multimodal feedback provided by the IVR application played complementary roles in preparing students for the final speech, including prompting immediate behavioral adjustment, raising awareness of areas for improvement, and anchoring instructor-led reflection. The repeated presence of a virtual audience also promoted students' preparedness for the live event, including forming strategies to manage audience-related anxiety. These effects were facilitated by an alignment between the IVR app's

affordances and the activity's object. Specifically, real-time prompts and summary analytics supplied needed information for personalized goal-setting, performance monitoring, and evaluation. And the realistic simulation and repeated practice opportunities helped the students build confidence.

Further, instructor feedback showed consistently higher uptake than app-generated feedback, indicating a stronger mediating role of the instructor. In most cases, learners frequently acted immediately on suggestions regarding pronunciation, tone, and lexical choice, and engaged in guided reflection to interpret app-generated analytics. These findings showed that reflection with some guidance was critical for interpreting and applying multimodal feedback. Especially when the app-generated feedback was unclear or inaccurate, students would defer to the instructor's guidance, as they consider the instructor's perspective and judgement as authoritative. This may also explain why students' rejection of instructor input was rare. These findings corroborate with earlier studies related to feedback effectiveness for developing public speaking skills, particularly the strong positive effect results from integrating instructor's feedback to self-assessment (Van Ginkel et al., 2019; Van Ginkel et al., 2020). Our study further suggests that the app-generated analytics enrich this coordination by surfacing delivery features that are otherwise hard to capture for human raters.

RQ3: What contradictions within the IVR activity system emerged around feedback utilization and public speaking practice, and how did students navigate these contradictions?

We identified contradictions on the primary (tensions within a specific node of the activity system), secondary (tensions between nodes within the system), and quaternary (between other activity systems). The primary tensions were concerned with limitations in the app's built-in functionalities, such as the virtual audience and the processing capability in the target language (Korean), which fell short of students' expectations of how they expected the tool to help. However, consistent with Engeström's (2001, 2014) view that contradictions can drive change, learners developed adaptive strategies. For instance, Student 3 redirected her attention to the timer to manage anxiety in lieu of audience reaction, then transferred the strategy to the live speech. Other contradictions emerged from the misfit between the tool, the task, and demands of another activity system (final speech event), with two design frictions surfaced. First, the Korean font rendering issues in the teleprompter had a minor impact as scripts were later memorized. Second, the lack of slide-deck integration during the practice affected final performance, as evidenced in the judges notes of misalignment between talk and visuals. This prompted the instructor's decision to explore other alternatives, including considering another IVR application, or changing the practice format by using a computer and recording device instead.

Compared to other studies of virtual worlds or VR using the AT framework (e.g., Couture-Matte, 2025; Ryder & Yamagata-Lynch, 2014), our studies identified relatively less contradictions. This might be partly due to our task design, learner support, and instructor preparation prior to students' engagement with the IVR application, which helped prevent some of the tensions and challenges identified in previous research. Clear task goals, one-on-one facilitation, learner maturity, pre-session orientation/warmups, and instructor mediation of app-generated analytics, all of which contributed to preempting common breakdowns. Where contradictions did arise, they either spurred innovation (strategy development, tool substitution) or dissipated as practices stabilized—illustrating AT's claim that contradictions are necessary but insufficient drives of change (Murphy & Rodriguez-Manzanarez, 2008; Engeström, 2014).

7 Implications

7.1 For L2 pedagogy for presentational skills

Based on instructional practices implemented and examined in this study, the following implications highlight actionable strategies for integrating IVR-supported multimodal feedback into L2 speaking

pedagogy. First, our findings show that technology-mediated multimodal feedback can meaningfully support L2 public speaking when it is interpretable and aligned with task goals. Particularly, task design and instruction are critical for the effective integration of IVR into L2 learning and instruction. Brief but targeted training on app-generated analytics (e.g., pacing, gaze, gesture, fillers) and dashboard with summarized patterns help students act on feedback with minimal expert mediation. To manage cognitive load in IVR, practice sessions can be staged and adaptive, with early sessions light and emphasizing one salient cue at a time, and reserve fuller analytics for post-task review; as familiarity grows, additional feedback types can be attended to.

Our findings further suggest that feedback modalities may be sequenced according to learners' proficiency levels. For lower-proficiency speakers, instructor feedback emphasizing verbal guidance and focused visual cues may support foundational public speaking skills, while more complex multimodal feedback can be provided as learners gain greater control and confidence. For higher-proficiency speakers, simultaneous multimodal feedback may foster deeper self-monitoring and metacognitive awareness.

Further, a deliberate division of labor can be effective. As the IVR application provides analytics on nonverbal performance, the instructor can target higher-order linguistic features. Our practice design has shown to be effective with a short reflection–rehearsal cycle in which students use the analytics to self-evaluate, set one actionable goal for the next attempt, and debrief with targeted prompts from the instructor. Repetitions in terms of rehearsal and exposure to a virtual audience also seem to be success factors, helping students build and transfer confidence for the live event.

Lastly, while this course was situated in a Korean Flagship context, the design principles are broadly applicable to instructional design for other L2 at advanced level, specifically, task-based pedagogy anchored in authentic presentations, explicit scaffolding of professional discourse, and the integration of IVR and multimodal analytics into iterative practice and core feedback loop.

7.2 For teacher professional learning with IVR and other technologies

While IVR offers substantial learning benefits, its impact hinges more on pedagogy than on the tool itself (Marinsky & Petersen, 2021). Effective teacher professional learning is essential for helping instructors interpret multimodal analytics, calibrate feedback priorities, and integrate app-generated data into pedagogically meaningful guidance. Instructors need opportunities to develop not only technical skills with IVR systems, but also pedagogical judgement about how to use the feedback based on task goals and learner needs. Such preparation enables instructors to interpret, prioritize, and pedagogically use multimodal feedback as part of a coherent instructional design. This places a demand on comprehensive teacher education—technical, pedagogical, and psychological—so instructors can integrate IVR purposefully and confidently (Cowie & Alizadeh, 2022; Kaplan-Rakowski & Gruber, 2023). In practice, teachers should anticipate mismatches between simulated IVR conditions and real-world performance contexts (e.g., Ignite Talks in our study) and plan targeted supplements, such as additional instructional materials, or alternative practice setups—so that skills transfer smoothly beyond the IVR settings.

7.3 For IVR application design

Drawing on our findings, we propose the following forward-looking recommendations for IVR application design based on observed affordances and limitations. First, to enhance feedback mechanisms, the virtual audience in IVR need to deliver immediate, nonverbal feedback tied to performance—e.g., dynamic gaze, posture shifts, murmurs, or applause that respond to pacing, clarity, and engagement. To support different target languages, it is also important to ensure full language compatibility across scripts, fonts, and interface elements (teleprompter, captions, menus), and smooth handling of the target language in analytics. Providing scaffolded configurations that can be differentiated

by learner readiness may also be highly valuable, for instance, having lighter cue density and simpler controls early on and gradually introducing richer analytics and more responsive audience behaviors as proficiency grows. Ideally, settings that allow cue prioritization (e.g., pacing first, then gaze/gesture) and customizable prompts (e.g., checkpoint-style rather than constant pop-ups) should be available so instructors can align the feedback stream with the task goals and learners' developmental stage (O'Connor & Worman, 2019).

8 Conclusions

Situated in an advanced Korean course, this study examined IVR-supported public speaking practice with multimodal feedback, using the Activity Theory framework. We examined students' response actions toward feedback from the app and the instructor evolved across sessions. Our findings suggested that different feedback sources played different roles in supporting learners' preparation for their final speech to a live audience. We also identified contradictions that arose from the mediational relationships among elements within and between activity systems. Taken together, the findings show that IVR can productively support L2 presentational skills with purposeful instructional design and learner support. Empirically, our study enriches understanding of the learning process in IVR by tracing how learners attend to, interpret, and act on multimodal analytics during rehearsal and reflection. Regarding learning design, we propose instructional principles for IVR integration—task–tool alignment, staged cueing to manage cognitive load, all of which are transferable beyond the Korean context to advanced L2 instruction more broadly.

Our study does have limitations, and the results should be interpreted with caution. Primarily, the practice window was short with relatively low frequency, and the implementation depended on instructor mediation to interpret analytics and guide reflection, which may have constrained fully autonomous engagement with real-time cues. Additionally, learners' improvements may not be attributed solely to IVR practice because they occurred within an ongoing course context; learning transfer from regular instruction and parallel supports may also have contributed to final performance. To further enhance our understanding of multimodal feedback delivered through IVR, future work can consider adopting other pedagogical models for task design, especially self-directed learning with no instructor supervision that incorporate pre-training, embedded explanations of metrics, and configurable feedback staging, then compare them with peer-facilitated and instructor-led formats across languages and presentation genres. Longer interventions with integrated presentation slides and more responsive audience behaviors can further test transfer of learning to live performance contexts.

Appendices

Appendix A. Coding Scheme

Code Category/Name	Description
INSTRUCTOR FEEDBACK UPTAKE	**How students respond to feedback from instructor after IVR practice**
FEED_ACCEPTED_DELAY	Learner applies instructor feedback in later sessions or final speech
FEED_ACCEPTED_IMM	Learner takes immediate action based on instructor feedback (e.g., adjusts gestures mid-practice)
FEED_REFLECTED	Learner discusses or evaluates instructor feedback
FEED_REJECTED	Learner chooses not to act on/accept instructor feedback, possibly with reason

OUTCOMES	**Observed, documented and/or perceived learning outcomes**
OUT_CONF_GAINED	Evidence of increased confidence (from surveys/interviews/ evaluation)
OUT_CONF_LOW	Negative evaluation on low confidence
OUT_NEGATIVE_REACT	Learner expresses negative reaction to feedback
OUT_SELF_NEG	Learner's self-evaluation of negative learning outcomes
OUT_SELF_POS	Learner's self-evaluation of overall learning outcomes
OUT_SKILL_IMPROVED	Observable performance gains (speech metrics, coach evals)
OUT_TIME_NEG	Negative evaluation on time management
TOOL-MEDIATED FEEDBACK	**Learner response to app-generated feedback (real-time and summary)**
APP_ACCEPTED_DELAY	Learner applies app feedback in later sessions or final speech
APP_IGNORED	Learner does not respond to app-generated feedback
APP_INSTR_GUIDED_POST	Instructor interprets/guides students through app generated feedback during review
APP_MISUNDERSTOOD	Learner misinterprets the app feedback, applies it incorrectly
APP_REFLECTED	Learner discusses or evaluates app-generated feedback
APP_REFLECTIVE_USE	Learner reflects on the role of the tool in their learning (from review, interviews or surveys)
TOOL DISCUSSION	**Learner discusses features of VR and their effect**
Tool negative	Learner discusses IVR features perceived negative
Tool positive	Learner discusses IVR features perceived positive

Appendix B. Rubric for evaluating final speech performance

평가기준 (Areas for Evaluation)	상세기준 (Description)
구성 (Organization) (10 points)	The presentation is organized and the ideas and images flow in a cohesive manner that is easily followed and understood with impeccable transition/signposting markers. Topic development followed by Introduction - Body - Conclusion organization can be clearly identified. Introductory and closing statement applies to the audience; Introductory statement presents flow of discussion and closing statement flawlessly summarizes, concludes, or restates topic of discussion.
내용 (Content) (10 points)	The presentation includes a clear purpose of the speech, which captures the audience's attention. Also, it contains appropriate content related to the provided theme, sustainability, showing coherency and incorporating specific examples. The presentation concludes with the important points and encourages future discussion.
문법 , 어휘 선정 , 표현 (Language) (10 points)	The presenter uses the right words, sentence structure, and correct grammar to communicate their ideas, showing effort to use rhetorical devices to enhance and emphasize ideas. Also, the presenter selects relevant information from numerous sources that can support their main ideas with specific facts, examples, and illustrations.

전달력 (Delivery) (10 points)	The presenter demonstrates presentation skills (culturally appropriate stance and pose, movement, gestures, facial expressions, eye contact, tone, volume, confidence, etc) to express their message and achieve their purpose of speech in a natural way, expressing the main points with a natural flow in terms of pace.
슬라이드 및 시각적 효과 (Visual Aids) (5 points)	The presenter uses visual aids (PowerPoint Slides) in a consistent in design in terms of the font and size of letters and color, embedding various types of aids (graph, table, chart, or diagram, etc) not only to express their main ideas, but also to maintain the attention of audience.
시간 관리 - 슬라이드 당 15 초 , 총 5 분 (Time management) (5 points)	The presenter is able to complete their speech within 5 minutes, synchronizing with their PowerPoint Slides. (Note that Ignite Talk is exactly 5 minutes long, and is comprised of 20 slides that auto-advance every 15 seconds)
태도 , 문화이해 (Attitude & Cultural Awareness) (5 points)	The presenter demonstrates motivation, determination, sincerity, passion, and maturity; demonstrates awareness of Korean culture (e.g., being humble, greeting). Also, the presenter is in a proper presentation attire, dressing professionally.

References

American Council on the Teaching of Foreign Languages [ACTFL]. (2012). *Performance descriptors for language learners.* https://www.actfl.org/uploads/files/general/ACTFLPerformance_Descriptors.pdf

Atkinson, D. (2010). Extended, embodied cognition and second language acquisition. *Applied Linguistics, 31*(5), 599–622. https://doi.org/10.1093/applin/amq009

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Beltrán-Palanques, V. (2024). Teaching elevator pitch presentations through a multimodal lens: Insights from ESP students' experiences. *TESOL Journal*, *15*(2), e769. https://doi.org/10.1002/tesj.769

Boetje, J., & van Ginkel, S. (2021). The added benefit of an extra practice session in virtual reality on the development of presentation skills: A randomized control trial. *Journal of Computer Assisted Learning, 37*(1), 253–264. https://doi.org/10.1111/jcal.12484

Carrell, L. J., & Menzel, K. E. (1997). The impact of preparation and motivation on learning performance. *Communication Education, 46*(4), 262–272. https://doi.org/10.1080/03634529709379100

Chun, D. M., Karimi, H., & Sañosa, D. J. (2022). Traveling by headset: Immersive VR for language learning. *CALICO Journal, 39*(2), 129–149. https://doi.org/10.1558/cj.21306

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge University Press.

Couture-Matte, R. (2025). Immersive virtual reality and language learning: activity theory perspectives. *Computer Assisted Language Learning, 1–30.* https://doi.org/10.1080/09588221.2025.2497494

Cowie, N., & Alizadeh, M. (2022). The affordances and challenges of virtual reality for language teaching. *International Journal of TESOL Studies, 4*(3), 50–65. https://doi.org/10.46451/ijts.2022.03.05

Creswell, J. W., & Poth, C. N. (2016). Qualitative inquiry and research design: Choosing among five approaches. Sage Publications.

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends, 66*(5), 810–824. https://doi.org/10.1007/s11528-022-00717-w

Ding, M. (2024). The impact of high-immersion virtual reality on EFL learners' foreign language speaking anxiety: A mixed-method approach. *ReCALL, 36*(3), 287–305. https://doi.org/10.1017/S0958344024000156

El Majidi, A., Janssen, D., & de Graaff, R. (2021). The effects of in-class debates on argumentation skills in second language education. *System, 101*, 102576. https://doi.org/10.1016/j.system.2021.102576

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research. Orienta-Konsultit Oy.*

Engeström, Y. (2014). Activity theory and learning at work. In *Tätigkeit-Aneignung-Bildung: Positionierungen zwischen Virtualität und Gegenständlichkeit* (pp. 67–96). Springer Fachmedien Wiesbaden.

Engeström Y. (2001). Activity theory as a framework for the study of organizational transformation. In: *Knowing in practice, February 2001, University of Trento. Available at:* http://blog.roodo.com/dpopen/e79f581b.doc.

Fincham, N. X., Rodríguez, J. C., & Tschudi, S. (2022). Workplace Simulations: Test-Driving Life in Professional Contexts. In Wang & Uber Grosse (Eds.), *Chinese for Business and Professionals in the Workplace: Reaching across Disciplines (1st ed.)* (pp. 238–255). Routledge. https://doi.org/10.4324/9781003014690

Garbati, J. F., & Mady, C. J. (2015). Oral skill development in second languages: A review in search of best practices. *Theory and Practice in Language Studies, 5*(9), 1763. http://dx.doi.org/10.17507/tpls.0509.01

Guichon, N., & Cohen, C. (2016). Multimodality and CALL. In F. Farr & L. Murray (Eds.), The Routledge handbook of language learning and technology (pp. 509–521). Routledge.

Jiang, L. G., & Hafner, C. (2024). Digital multimodal composing in L2 classrooms: A research agenda. Language Teaching, 58, 1–19. doi:10.1017/S0261444824000107

Kaplan-Rakowski, R., & Gruber, A. (2023). The impact of high-immersion virtual reality on foreign language anxiety. *Smart Learning Environments, 10*(46). https://doi.org/10.1186/s40561-023-00263-9

Kaplan-Rakowski, R. (2025). Virtual Reality–Assisted Language Learning (VRALL). In The Palgrave Encyclopedia of Computer-Assisted Language Learning (pp. 1–7). Springer Nature Switzerland.

Khonamri, F., Králik, R., Vítecková, M., & Petrikovicová, L. (2021). Self-Assessment and EFL Literature Students' Oral Reproduction of Short Stories. *European Journal of Contemporary Education, 10*(1), 77–88. https://doi.org/10.13187/ejced.2021.1.77

King, P. E., Young, M. J., & Behnke, R. R. (2000). Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions. *Communication Education, 49*(4), 365–374. https://doi.org/10.1080/03634520009379224

Lantolf, J. P. (2006). Language emergence: Implications for applied linguistics—A sociocultural perspective. *Applied Linguistics, 27*(4), 717–728. https://doi.org/10.1093/applin/aml034

Lee, S. M., Wang, X., Park, I., & Lestiono, R. (2024). It feels so real! Situated authentic language learning in immersive virtual reality. *Education and Information Technologies, 29*(18), 24023–24045. https://doi.org/10.1007/s10639-024-12807-8

Le Vo, T. H. (2022). Online simulated workplace tasks to enhance business English learning. *Journal of Asian Business and Economic Studies, 29*(3), 205–221. https://doi.org/10.1108/JABES-06-2020-0058

Liu, T., & Aryadoust, V. (2024). Orchestrating teacher, peer, and self-feedback to enhance learners' cognitive, behavioral, and emotional engagement and public speaking competence. *Behavioral Sciences, 1*4(8), 725. https://doi.org/10.3390/bs14080725

Lyster, R., & Saito, K. (2010). *Oral feedback in classroom SLA: A meta-analysis. Studies in second language acquisition, 32*(2), 265–302. https://doi.org/10.1017/S0272263109990520

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational psychology review, 33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In Mayer, R. E. (Ed.), The Cambridge handbook of multimedia learning (pp. 31–48). Cambridge university press.

Miskam, N., & Saidalvi, A. (2020). Using video technology to improve oral presentation skills among undergraduate students: A systematic literature review. *International Journal of Psychosocial Rehabilitation*, *24*(5), 5280–5291. https://doi.org/10.37200/IJPR/V24I5/PR2020235

Murphy, E., & Rodriguez-Manzanarez, M. A. (2008). Using activity theory and its principle of contradictions to guide research in educational technology. *Australasian Journal of Educational Technology, 24*(4), 442–457.

Nocchi, S. (2018). Foreign language teaching and learning in virtual worlds: The construct of affordance. *In Virtual worlds: concepts, applications and future directions, 169–200.*

Ochoa, X., & Zhao, H. (2024). OpenOPAF: An Open-Source Multimodal System for Automated Feedback for Oral Presentations. *Journal of Learning Analytics, 11*(3), 224–248. https://doi.org/10.18608/jla.2024.8411

O'Connor, E. A., & Worman, T. (2019). Designing for interactivity, while scaffolding student entry, within immersive virtual reality environments. *Journal of Educational Technology Systems, 47*(3), 292–317. https://doi.org/10.1177/0047239518817545

Perez, M. M. (2020). *Multimodal input in SLA research. Studies in Second Language Acquisition, 42*(3), 653–663. https://doi.org/10.1017/S0272263120000145

Ryder, L., & Yamagata-Lynch, L. (2014). Understanding tensions: Activity systems analysis of transpacific collaboration. *CALICO Journal, 31*(2), 201–220. https://doi.org/10.11139/cj.31.2.201-220

Seuling, P. D., Czernin, N. S., & Schiele, M. A. (2024). Virtual Reality exposure therapy in the treatment of public speaking anxiety and social anxiety disorder. *Neuroscience applied, 3*, 104074. https://doi.org/10.1016/j.nsa.2024.104074

Sun, Y. C., & Yang, F. Y. (2013). I help, therefore, I learn: service learning on Web 2.0 in an EFL speaking class. *Computer Assisted Language Learning, 28*(3), 202–219. https://doi.org/10.1080/09588221.2013.818555

Swender, E., Conrad, D. J., & Vicars, R. (2012). ACTFL proficiency guidelines 2012. *American Council on the Teaching of Foreign Languages.*

Tatzl, D. (2017). An intensive presentations course in English for aeronautical engineering students using cyclic video recordings. *Language Learning in Higher Education, 7*(2), 275–309. https://doi.org/10.1515/cercles-2017-0011

Thorne, S. L. (2003). Artifacts and cultures-of-use in intercultural communication. *Language Learning & Technology, 7*(2), 38–67.

Thrasher, T. (2022). The impact of virtual reality on L2 French learners' language anxiety and oral comprehensibility: An exploratory study. *CALICO Journal, 39*(2). https://doi.org/10.1558/cj.42198

Tschirner, E. (2001). Language acquisition in the classroom: The role of digital video. *Computer assisted language learning, 14*(3-4), 305–319.

Van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education. *Educational Research Review, 14,* 62–80. https://doi.org/10.1016/j.edurev.2015.02.002

Van Ginkel, S., Gulikers, J., Biemans, H., Noroozi, O., Roozen, M., Bos, T., ... & Mulder, M. (2019). Fostering oral presentation competence through a virtual reality-based task for delivering feedback. *Computers & Education, 134*, 78–97. https://doi.org/10.1016/j.compedu.2019.02.006

Van Ginkel, S., Ruiz, D., Mononen, A., Karaman, C., De Keijzer, A., & Sitthiworachart, J. (2020). The impact of computer-mediated immediate feedback on developing oral presentation skills: An exploratory study in virtual reality. *Journal of Computer Assisted Learning, 36*(3), 412–422. https://doi.org/10.1111/jcal.12424

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Wang, B., Yu, S. & Teo, T. (2018). Experienced EFL teachers' beliefs about feedback on student oral presentations. Asian. J. Second. Foreign. Lang. Educ. 3, 12. https://doi.org/10.1186/s40862-018-0053-3

Weimer, M (2013). *Learner-centered teaching: Five key changes to practice.* John Wiley & Sons.

Wilson, M. Six views of embodied cognition. *Psychonomic Bulletin & Review, 9, 625–636* (2002). https://doi.org/10.3758/BF03196322

Wörtwein, T., Morency, L. P., & Scherer, S. (2015, September). Automatic assessment and analysis of public speaking anxiety: A virtual audience case study. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 187–193). IEEE.

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., ... & Yu, L. F. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality, 2.* https://doi.org/10.3389/frvir.2021.645153

Yoon, S. S., & Seo, J. M. (2024). Korean language educators' perceptions on the language of instruction in Korean as a Foreign Language classes in North America. *Journal of the National Council of Less Commonly Taught Languages, 36*(S2), 167–216.

Naiyi Xie Fincham holds a PhD in Educational Psychology and Educational Technology from the Michigan State University, USA. Currently she is a Specialist Faculty in Learning Design at the Center for Language & Technology, University of Hawai'i at Mānoa. Her background lies at the intersection of second language pedagogy, computer-assisted language learning, learner agency, and learning design. Her research and professional work focus on innovative, technology-enhanced pedagogies and fostering learners' self-regulated learning through intentional design and targeted scaffolding. Her recent projects examine how generative AI and mixed-reality environments can support language development.

Leeseul Park holds a PhD in East Asian Languages and Literatures from the University of Hawai'i at Mānoa, USA. Currently she is the Korean Language Program Director, Department of Comparative Literature and Intercultural Studies at the University of Georgia, USA. Her research focuses on Korean linguistics, second language acquisition, and language assessment, with particular attention to CAF-based proficiency measures, instructional materials development, and technology-enhanced Korean language pedagogy.