ISSN: 2473-4901
v12 n6706

Practice, Presence, and Performance: A Mixed-Methods Study of Repeated VR Roleplay in Undergraduate Courses

Kevin Mentzer
Kevin.Mentzer@Nichols.edu

Jason A. Price
Jason.Price@Nichols.edu

Andrea Becker
Andrea.Becker@Nichols.edu

Beth Gionfriddo
Beth.Gionfriddo@Nichols.edu

Nichols College
Dudley, MA, 01571, USA

Abstract

This mixed-methods study examines repeated immersive virtual reality (IVR) roleplay as a tool for developing speaking performance in higher education. Undergraduate students completed VR-based speaking scenarios in two instructional contexts: a first-year course (LEAD 120) and a junior-level seminar course (PDS). Quantitative analyses compared first and second attempts on repeated assignments using platform-generated performance metrics, including overall, content, and delivery scores, as well as duration, words per minute, filler words, listenability, eye contact, and body language. Qualitative student reflections were analyzed thematically to better understand how learners experienced the affordances and constraints of VR-based speaking practice. Across repeated roleplays, second attempts were generally associated with improvement on several performance metrics, particularly in the first-year cohort. LEAD 120 students showed clearer gains than PDS students, suggesting that repeated VR practice with reflective AI coaching may be especially beneficial for less experienced speakers who have greater room for short-term growth. Qualitative findings indicate that students in both cohorts often valued the experience for practice, feedback, and self-awareness, although they differed in how they described its primary benefits. The study contributes evidence that repeated VR roleplay and reflective AI coaching can function as a meaningful form of experiential speaking practice while also highlighting the importance of learner experience, prior preparation, and task design in shaping outcomes.

Keywords: immersive virtual reality (IVR), communication skills, roleplay, AI coaching, experiential learning, higher education

ISSN: 2473-4901
v12 n6706

ISSN: 2473-4901
v12 n6706

Practice, Presence, and Performance: A Mixed-Methods Study of Repeated VR Roleplay in Undergraduate Courses

Kevin Mentzer, Jason Price, Andrea Becker and Beth Gionfriddo

1. Introduction

Virtual reality (VR) continues to attract attention as an instructional technology because it can simulate authentic environments in ways that deepen attention, participation, and experiential engagement. Across educational settings, immersive virtual reality (IVR) has been used to support a wide range of outcomes, including technical skill development, spatial understanding, and communication practice. Prior work has also extended VR into higher education business and communication contexts, where simulated environments can provide students with opportunities to rehearse difficult interpersonal situations in low-risk settings. Research in business and communication education suggests that VR can increase student engagement, confidence, and self-awareness, while also revealing practical implementation constraints such as setup time, usability friction, and limited instructor control over scenario design (Mentzer et al., 2025; Patterson et al., 2025).

At the same time, the educational value of VR roleplay likely depends on more than novelty alone. Prior work suggests that immersive technologies are more likely to support learning when learners experience both presence and meaningful agency, and when instructional design manages cognitive load, task alignment, and learner processing rather than relying on immersion alone (Makransky & Petersen, 2021; Mayer et al., 2023). The Cognitive Affective Model of Immersive Learning (CAMIL) is useful here because it positions presence and agency as central to shaping related outcomes such as interest, motivation, self-efficacy, embodiment, cognitive load, and self-regulation (Makransky & Petersen, 2021).

The present study builds on that earlier work by shifting from a primarily qualitative exploration of learner experience to a mixed-methods analysis of repeated VR speaking performance followed by reflective AI coaching. Rather than asking only how students perceived IVR, this study asks what changed when students repeated the same VR roleplay assignment and how student reflections may help explain those changes.

This question is especially relevant across different student populations. In the current dataset, first-year students in LEAD 120 completed a structured series of VR roleplays designed to include repeated attempts, while junior-level students in PDS repeated a professional interview scenario. These two instructional contexts provide an opportunity not only to examine overall first-to-second-attempt change, but also to consider whether repeated VR practice functions differently for students at different stages of their academic and professional development.

This study addresses three research questions:

> **RQ1:** How do students' platform-generated speaking performance metrics change across repeated attempts in AI-supported VR roleplay?
> **RQ2:** How do patterns of speaking performance change vary across student cohorts and instructional contexts?
> **RQ3:** How do student reflections help explain the perceived value and limitations of repeated VR roleplay and AI-guided coaching?

To address these questions, this study uses an explanatory mixed-methods design. Quantitative analyses compare first and second attempts using platform-generated performance metrics, while qualitative reflections are used to interpret patterns related to confidence, realism, feedback, pacing, and repetition.

ISSN: 2473-4901
v12 n6706

2. Literature Review

Immersive virtual reality (IVR) has gained increasing attention as an instructional technology because it can place learners in simulated environments that support experiential, contextualized, and practice-based learning. Early reviews identified VR's potential to provide authentic, interactive learning experiences while also noting limitations related to cost, accessibility, and instructional design (Freina & Ott, 2015).As VR hardware and software have matured, more recent studies suggest that instructional design remains the primary determinant of learning outcomes, with engagement and learning benefits depending on how immersive experiences are structured rather than on immersion alone (Cabrera-Duffaut et al., 2024; Kaimara et al., 2022).

Recent work reinforces this pattern in business and communication education. Mentzer et al. (2025) found that VR-enhanced business learning increased student engagement, while implementation challenges such as setup time and limited content control constrained the experience. Patterson et al. (2025) similarly showed that VR can support confidence and self-awareness in speaking instruction, suggesting that immersive environments may be especially useful when they structure rehearsal, reflection, and feedback around communication tasks.

Speaking-intensive learning environments are a particularly strong fit for IVR because effective oral communication depends on practice, context, and repeated exposure to realistic social situations. Traditional classroom instruction can support speaking development through presentations, discussions, and feedback, but it may provide fewer opportunities for students to rehearse the same high-stakes interpersonal situation more than once under comparable conditions. By contrast, VR roleplay environments can simulate difficult conversations, collaborative problem-solving scenarios, and interviews in a low-risk setting while giving learners the chance to repeat the activity and reflect on their performance. This repeated-practice dimension is especially important for the present study, which focuses not only on learner perceptions of VR but on measurable first-to-second-attempt change.

The Cognitive Affective Model of Immersive Learning (CAMIL) explains immersive learning as an interaction between features of the technology and learner-centered psychological processes. Presence refers to the learner's sense of "being there," while agency reflects the perceived ability to influence one's actions and environment. More broadly, agency has long been recognized as a fundamental construct in cognitive psychology describing the subjective experience of controlling one's own actions and their consequences (Moore & Fletcher, 2012). Within CAMIL, presence and agency shape motivation, self-efficacy, cognitive load, self-regulation, and ultimately learning outcomes (Makransky & Petersen, 2021). CAMIL suggests that immersive learning is most productive when these elements are aligned in ways that promote meaningful engagement rather than distraction or overload.

Recent work also shows that CAMIL can be adapted productively to specific educational settings. In a qualitative study of immersive VR in middle school Spanish, Loughlin and Mentzer (2026) applied an adapted CAMIL framework and found that IVR supported engagement, perceived learning, and confidence when presence was paired with meaningful learner agency and when implementation reduced friction and supported autonomy. At the same time, that study found that technical barriers and constrained control could weaken the potential benefits of immersion. This is especially relevant to the present study because student reflections in both LEAD 120 and PDS likewise address realism, confidence, feedback, usability, and the limits of AI-mediated interaction.

Immersive practice represents only one component of effective learning. Research on intelligent tutoring systems has consistently demonstrated that timely, individualized feedback promotes reflection, self-regulated learning, and improved performance (Ma et al., 2014). Most recently, conversational AI has expanded these capabilities by providing learners with immediate formative coaching following authentic performance tasks. Rather than simply evaluating performance, AI-guided feedback can prompt learners to identify strengths, recognize weaknesses, and develop strategies for subsequent attempts. The present study builds on this by combining immersive VR roleplay with AI-guided coaching and repeated guidance.

For the present study, CAMIL is best treated as a light interpretive framework rather than as a rigid explanatory model for every quantitative outcome. The model helps explain why immersive experiences

ISSN: 2473-4901
v12 n6706

may feel engaging, realistic, and confidence-building even when measurable performance changes are uneven across tasks or student groups. However, the current intervention also includes a feature that pushes beyond a simple CAMIL application: Rather than viewing the intervention as immersive roleplay alone, the present study conceptualizes it as an instructional cycle consisting of immersive practice, AI-guided formative feedback, learner reflection, and repeated performance.

3. Methodology

This study used an explanatory mixed-methods design. Quantitative analyses examined first-to-second-attempt change using platform-generated performance metrics, while qualitative reflections were analyzed to interpret how students conceptualized repetition, realism, confidence, pacing, and feedback.

The study was conducted at a New England college and focused on students enrolled in two primary course contexts. The first was LEAD 120, a required first-year course taken by incoming first-year and transfer students. The second was PDS, a junior-level professional-development seminar course. Additional VR roleplay records from other courses/contexts were also available in the dataset and were included in pooled descriptive and overall repeat-attempt analyses, but they were not treated as focal groups for course-specific interpretation.

Participants in the LEAD 120 cohort were primarily traditional-aged undergraduate students enrolled in a first-year leadership course at a four-year institution. Participants in the PDS cohort were primarily traditional-aged undergraduate students enrolled in a junior-level course in their third year of study at the same institution. Both LEAD 120 and PDS participants represented a broad range of academic majors offered by the institution.

PDS participants were primarily traditional college-aged undergraduate students enrolled in a junior-level course at a four-year institution. The majority were in their third year of study covering all offered majors at the institution.

The LEAD 120 implementation was designed around a structured sequence of VR activities. Students first completed a familiarization task intended to orient them to the technology and format. They were then expected to complete Activity 1 twice during the first class session, followed later by Activity 2 twice during a second class session. Students who followed the intended protocol therefore generated five total records across the sequence. In the PDS context, students completed the "Interview on Uploaded Resume" scenario twice. These curricular differences allowed the study to compare repeated practice across a novice-oriented first-year context and a more advanced professional-development context.

Each VR activity included both a simulated roleplay and a follow-up feedback session. In the initial phase, students interacted with an AI-driven avatar in a scenario tied to a specific communication context, such as a difficult conversation or interview. After completing the scenario, students entered a second virtual meeting space in which a different AI avatar reviewed their performance. This follow-up exchange was interactive rather than limited to a static score report: the avatar identified strengths and weaknesses, asked students reflective questions about what they believed went well and discussed areas for improvement. The instructional design therefore combined immersive practice with immediate AI-guided coaching and reflection.

The system also generated session-level performance reports for each activity. These reports included quantitative indicators including overall score, content score, delivery score, duration, words per minute, filler words, listenability, eye contact, and body language. Listenability is a platform-generated measure of vocabulary, sentence structure, and clarity. Because the platform emphasized simulated interpersonal speaking scenarios, it functioned as a form of guided experiential practice rather than as a static content-delivery system.

Quantitative data consisted of exported HTML roleplay reports collected from platform archives. These files were flattened into a structured dataset in spreadsheet format, with one row per roleplay record. Variables extracted from the reports included student name, email, roleplay type, session date and time, and all available numeric performance measures. To identify repeated attempts, records were grouped

ISSN: 2473-4901
v12 n6706

by student and roleplay type. For the main paired analyses, only cases with at least two attempts on the same assignment were retained. Attempts were then ordered chronologically so that the earliest valid attempt served as Attempt 1 and the next valid attempt served as Attempt 2. For students with more than two attempts on the same roleplay, the primary paired analyses focused on the first two attempts in order to preserve a consistent comparison across tasks.

Descriptive statistics were computed for first and second attempts, and paired t-tests were used to evaluate whether average differences were statistically significant. Raw p-values were reported for consistency with common practice in the target outlet. Analyses were conducted at three levels: overall pooled repeated roleplays, focal LEAD 120 repeated assignments, and the focal PDS repeated assignment.

The qualitative component consisted of student reflections gathered about their experiences using VR roleplay in both LEAD 120 and PDS. These responses were used to identify themes related to realism, confidence, coaching and feedback, perceived improvement, technical friction, and the broader developmental value of VR-based speaking practice. Qualitative data were analyzed thematically using a hybrid deductive-inductive coding approach informed by CAMIL and by emergent themes in the responses. Initial codes addressed perceived value, realism, confidence, coaching and feedback, perceived improvement, communication self-awareness, transfer to real-world situations, technical or design friction, and preferences for human interaction. These qualitative findings were then used to interpret and contextualize the quantitative results.

4. Results

The final dataset consisted of roleplay records exported from the VR platform and flattened into spreadsheet form, with one row per session. For the repeated-attempt analyses, records were grouped by student and roleplay type, and only cases with at least two attempts on the same assignment were retained. Across all assignments, the pooled repeated-attempt sample included 222 student-by-roleplay pairs. However, sample size varied slightly by metric because some exported session reports did not contain complete values for every platform-generated measure. Full pooled and activity-level statistical tables are included in Appendix A and Appendix B.

Across all repeated VR roleplay assignments, second attempts were generally associated with improved performance on several metrics. In the pooled sample, overall score increased from 6.43 on Attempt 1 to 6.71 on Attempt 2 ($p < .0001$), content score increased from 6.38 to 6.68 ($p < .0001$), and delivery score increased from 6.85 to 6.99 ($p = .0023$). Students also showed higher eye contact ($p < .0001$), used fewer filler words ($p = .0420$), exhibited faster speaking rates ($p < .0001$), and had shorter duration ($p = .0013$). At the same time, listenability declined from 0.772 to 0.746 ($p < .0001$), suggesting that repeated roleplay may have increased efficiency and fluency while sometimes reducing clarity.

The clearest course-specific effects emerged in LEAD 120, the first-year course in which students completed a structured sequence of VR roleplays across two class sessions. During the first session, students completed a familiarization task intended to orient them to the headset hardware and the VirtualSpeech environment, followed by two attempts at Activity 1. In the second session, students completed two attempts at Activity 2. Because the familiarization task functioned primarily as an onboarding exercise rather than as a core instructional outcome, the repeated-attempt analyses were limited to Activity 1 and Activity 2.

For Activity 1, the results showed a generally favorable but mixed pattern of change. Overall and content scores increased modestly, but neither change reached conventional statistical significance. Delivery score improved significantly ($p = .0435$), as did eye contact ($p = .0066$). Students also spoke substantially faster on the second attempt ($p < .0001$) and completed the activity more quickly ($p = .0023$). Listenability declined significantly. Taken together, these findings suggest that early repeated practice may have helped students become more efficient and somewhat stronger in delivery and eye contact even though the larger scoring gains were still limited at this stage.

 ISSN: 2473-4901
v12 n6706

Activity 2 showed a stronger and more coherent pattern of improvement than Activity 1. Overall score increased from 6.59 to 6.94 ($p = .0008$), and content score increased from 6.55 to 6.94 ($p = .0006$). Students also became faster and more concise on their second attempt, with words per minute increasing from 127.3 to 147.2 ($p = .0001$) and duration decreasing from 337.1 seconds to 297.3 seconds ($p = .0051$). Eye contact improved ($p = .0286$), and body language improved ($p = .0198$). However, Listenability declined from 81.9% to 77.8% ($p < .0001$). This suggests that students' second attempts may have become faster and more efficient while not always becoming easier for an audience to process.

The PDS analysis focused on the Interview on Uploaded Resume assignment. In contrast to LEAD 120, the PDS results showed a flatter pattern of change. Overall and content scores increased modestly, but neither change reached statistical significance. Eye contact, listenability, words per minute, filler words and duration also showed limited movement. Overall, the PDS findings suggest that repeated VR roleplay may have produced directional improvement for some measures, but the magnitude of change was smaller and less consistent than in LEAD 120.

Table 1 presents a compact summary of the focal quantitative findings. The full pooled and activity-level statistical results are provided in Appendix A and Appendix B.

Context	Overall score change	Content score change	WPM change	Listenability change	Duration change
LEAD 120 Activity 1	+0.25 (ns)	+0.25 (ns)	+24*	-2.3%*	-49 sec*
LEAD 120 Activity 2	+0.36*	+0.4*	+19.9*	-4%*	-39.8 sec*
PDS Interview	+0.24 (ns)	+0.27 (ns)	+0.8 (ns)	+0.5% (ns)	+3.2 sec (ns)

Note. Asterisks indicate raw $p < .05$.

Table 1. Key first-to-second attempt changes by instructional context

4.1 Qualitative Findings

Qualitative reflections were analyzed to better understand how students experienced VR roleplay and how those experiences help explain the quantitative patterns observed across course contexts. Although both cohorts generally viewed the activities as useful, the nature of that perceived value differed somewhat between LEAD 120 and PDS. In LEAD 120, students more often emphasized growth through repeated practice, coaching, and application to teamwork and difficult conversations. In PDS, students more often framed the experience in terms of interview preparation, confidence building, and feedback, while also expressing greater skepticism about whether AI-based roleplay could approximate a live human interview.

For LEAD 120 students, two themes were especially prominent. The first was the value of coaching and repeated practice. Students described the post-scenario coach as one of the most useful parts of the experience, with one student writing, "I really liked the feedback from the coach. I found that to be the most beneficial piece of the activity." Longer reflections also linked improvement from the first session to the second to prior feedback. For example, one student explained, "By the second session, I learned to slow down, ask more helpful questions, and think about perspectives." These comments support the interpretation that the instructional value of the LEAD 120 activities came from the combination of practice, coaching, and a second opportunity to apply what had been learned.

The second major LEAD 120 theme was transfer to real-world communication, teamwork, and leadership. Students frequently described the activities as useful preparation for future classroom, workplace, and interpersonal situations. One student wrote that the experience "helped me become a better communicator and teammate" and overall made them feel "more ready to handle tough conversations and work with all kinds of people." Another reflected that the VR exercises helped them understand "how important communication and respect are to having good group work experiences." Students also became more aware of specific communication habits such as rushing, weak eye contact, filler words, unclear explanations, and tone. These responses suggest that LEAD 120 students often saw the experience as relevant not only to the immediate activity, but to broader leadership and collaboration challenges.

ISSN: 2473-4901
v12 n6706

At the same time, LEAD 120 students did not respond uncritically to the technology. Some described the simulations as realistic or helpful, but others emphasized their artificial qualities. In the short-form responses, one student commented, "I wasn't a big fan because it was very robotic and wasn't how an actual conversation would go," while another wrote, "I feel like you could tell the conversation was fake." These comments suggest that even among students who found the activities useful, realism and conversational naturalness remained important constraints.

For PDS students, two somewhat different themes stood out. The first was confidence and lower-stakes preparation. Many students described the VR interview as useful because it allowed them to practice in a setting that felt realistic enough to matter, but less intimidating than a live interview. One student explained that "the interviewer being virtual made me less afraid of messing up because it was not a real person that could judge me," while another wrote that the experience was helpful because "I wasn't as nervous going into it." In this cohort, the simulation appears to have functioned as a lower-pressure environment in which students could rehearse interview behavior and become more comfortable with the experience of being evaluated.

The second major PDS theme was the value of feedback and self-awareness, paired with reservations about realism relative to live human interviews. Students frequently emphasized the usefulness of the platform's feedback. One student noted that the system "docked me for speaking too fast and not using the STAR method correctly," while another said the experience helped them realize they "should be more concise" in a professional setting. A particularly notable comment came from a student who reported that "on my 2nd interview, I scored higher… as a result of me implementing the recommendations from the previous coaching." At the same time, many PDS students were careful to distinguish the VR interview from a live human interaction. One student wrote that "practicing with real interviewers will be more valuable," while another remarked that the AI interviewers were "very monotone" and "did not match how I think I did at all." These comments suggest that PDS students often saw the VR interview as useful preparation, but not as a replacement for real interview practice.

Although student perceptions were largely positive, the qualitative data suggests that current VR interview technologies may not fully replicate the complexity of human-led interviews. Participants cited monotone interviewer delivery, repetitive questioning, limited assessment of body language, and occasional technical malfunctions as barriers to realism. These findings indicate that while VR can serve as an effective supplementary practice tool, it should not be viewed as a complete replacement for live mock interviews.

Taken together, the qualitative findings clarify the quantitative pattern. LEAD 120 students more often described the VR activities as a growth-oriented cycle of practice, coaching, and application that supported teamwork and difficult conversations, which fits the stronger first-to-second-attempt gains seen in that cohort. PDS students, by contrast, more often emphasized confidence, feedback, and professional preparation, which helps explain why the interview simulation may have felt educationally valuable even though the measurable short-term gains were more modest.

5. Discussion

The findings suggest that the instructional value of VR roleplay lies less in immersion alone than in the cycle of simulated performance, AI-guided coaching, reflection, and repeated practice. Across the pooled sample, second attempts generally improved, but the decline in listenability suggests an important fluency-versus-clarity tradeoff.

The clearest quantitative gains appeared in LEAD 120, particularly in Activity 2. In that setting, students demonstrated broader quantitative improvement than the PDS cohort. This pattern suggests that repeated VR roleplay may be especially useful for earlier-stage learners who are still developing confidence and basic conversational control in structured interpersonal situations. The LEAD 120 qualitative findings reinforce this interpretation. Students frequently described the activities as useful preparation for teamwork, difficult conversations, and broader leadership situations, and many explicitly linked their improvement to the post-scenario coaching process. Their reflections suggest that the instructional value of the experience came not only from exposure to realistic scenarios, but from being prompted to reflect on what had gone well, what had not, and how to adjust on the next attempt.

ISSN: 2473-4901
v12 n6706

By contrast, the PDS cohort showed a flatter and less consistent quantitative pattern. None of the focal PDS metrics changed significantly from the first to the second attempt. However, the qualitative findings complicate any simple reading of these results as evidence that the activity was ineffective. PDS students often described the VR interview as useful for confidence building, lower-stakes preparation, and self-awareness about interview technique. They also repeatedly emphasized the value of the feedback process, even when they questioned whether the simulation fully approximated a live human interview. This suggests that the educational value of the PDS experience may have been more affective and reflective than immediately performance-based. In other words, the interview simulation may have helped students feel more prepared and more aware of their weaknesses even when those gains did not translate into large short-term score increases between attempts.

This contrast between LEAD 120 and PDS is one of the most important findings of the study. One plausible interpretation is that repeated VR roleplay may be especially beneficial for less experienced students who have more room for visible short-term growth. First-year students may benefit from the combination of structured practice, reduced stakes, and immediate coaching because they are still learning how to manage tone, pacing, eye contact, explanation, and conflict navigation in professional or semi-professional settings. Junior-level students in PDS, by contrast, may already possess more developed communication habits and may therefore require more challenging scenarios, richer human feedback, or a longer intervention to produce comparable measurable changes. This interpretation is consistent with the paper's broader argument that the same VR system may function differently across student developmental stages and instructional contexts.

Another key finding concerns the role of coaching and reflection. The current intervention was not simply a repeated roleplay exercise; each activity also included an AI-guided feedback conversation that reviewed strengths, weaknesses, and areas for improvement. Our qualitative findings suggest that students often perceived this coaching component as one of the most valuable elements of the experience. This indicates that the mechanism of learning may be better understood as a cycle of performance, coaching, reflection, and revised performance rather than as repetition alone. This interpretation aligns with prior work suggesting that immersive learning is most effective when practice is paired with meaningful learner processing and feedback. (Makransky & Petersen, 2021; Mayer et al., 2023).

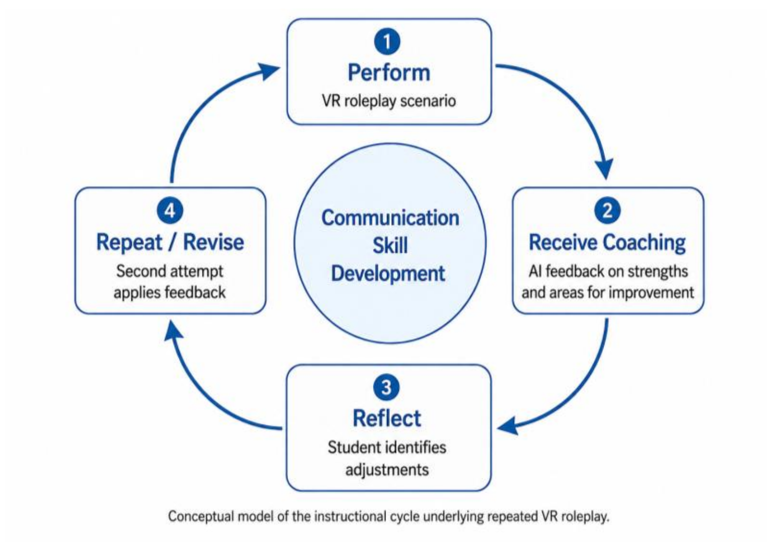

Figure 1. Conceptual Model of Practice and Reflection Cycle

Our findings also point to an important nuance in how improvement should be interpreted. On several measures, especially in LEAD 120, students became faster and more efficient on the second attempt. However, these gains were accompanied in some cases by lower listenability. Because listenability reflects vocabulary, sentence structure, and clarity, the persistent decline in this score complicates any simple interpretation of faster second attempts as better communication. Students may have become more fluent, more decisive, or less hesitant, but their responses may not always have become easier for an audience to digest. This fluency-versus-clarity tradeoff indicates that instructors using VR for speaking development may need to explicitly coach students not only toward confidence and pacing, but also toward clear, listenable delivery. In future implementations, this could mean placing more emphasis on reflective discussion of pacing, pauses, and audience comprehension rather than treating increased speed as inherently desirable.

ISSN: 2473-4901
v12 n6706

The qualitative findings also show that students valued the realism achieved through VR, but only up to a point. Students in both cohorts often described the simulations as useful and sometimes realistic, but many also noted that the AI interaction could feel robotic, repetitive, monotone, or unlike a genuine human exchange. This tension echoes CAMIL-based work suggesting that immersive environments can support engagement and perceived learning, but that these benefits depend on the balance between presence, agency, and implementation quality. In the current study, realism appears to have been "good enough" to make the activities useful for practice, but not good enough to eliminate comparisons with live human interaction. This is especially evident in the PDS interview context, where students were more likely to evaluate the simulation against real interviews and mock interviews with actual people. As a result, the present findings suggest that AI-mediated roleplay may work especially well as a preparatory or supplementary experience but should not be assumed to replace live interpersonal practice.

6. Limitations

This study has several limitations. First, it was conducted in a single institutional setting and across two course contexts that differed in both student population and assignment design. As a result, the LEAD 120 and PDS comparisons are informative but should not be treated as controlled experimental contrasts. Second, the study relied on platform-generated performance metrics rather than independent human ratings of speaking quality. These measures were useful for detecting patterns of change across attempts, but they should be interpreted as indicators of performance rather than definitive measures of communication competence. Third, prior speaking experience, communication confidence, and classroom preparation were not measured independently, limiting the ability to determine whether stronger LEAD 120 gains were due to developmental stage, task fit, baseline differences, or some combination of these factors. Fourth, qualitative reflections were self-reported and therefore subject to selectivity, recall bias, and differences in response depth. Finally, although the roleplay plus coaching structure appears promising, this study cannot isolate whether improvement was driven primarily by repetition, AI-guided feedback, reduced anxiety, or the interaction of these elements.

7. Conclusions and Future Work

This study suggests several directions for future work. Future research should examine repeated VR roleplay over longer time horizons and incorporate independent human ratings or rubric-based assessments alongside platform metrics. Comparative studies could also test whether novice learners benefit more than advanced learners from the same intervention or whether more experienced students require more complex scenarios, additional cycles of feedback, or live coaching to produce comparable gains. It would also be useful to compare AI-mediated coaching with human coaching or peer debriefing to better understand what kinds of reflection produce the strongest improvement.

Overall, this study suggests that repeated VR roleplay combined with reflective AI coaching can function as a meaningful form of experiential speaking practice in higher education. The strongest measurable gains appeared in the first-year LEAD 120 setting, while the PDS findings suggest that value may also appear in forms such as confidence building, preparation, and self-awareness, even when short-term score gains are modest. For instructors, the results support the use of VR not as a replacement for human interaction, but as a structured practice environment that can help students rehearse, reflect, and refine communication skills before higher-stakes real-world encounters.

8. Acknowledgments

Generative AI was used to assist with language refinement and readability improvement during preparation of the final manuscript draft. The authors reviewed and revised all content and take full responsibility for the originality, accuracy, and integrity of the submission.

The contents of this research paper were developed under the Department of Education grant #P116Z23

 ISSN: 2473-4901
v12 n6706

9. References

Cabrera-Duffaut, A., Pinto-Llorente, A. M., & Iglesias-Rodríguez, A. (2024). Immersive learning platforms: Analyzing virtual reality contribution to competence development in higher education—a systematic literature review. Frontiers in Education, 9. https://doi.org/10.3389/feduc.2024.1391560

Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. The International Scientific Conference eLearning and Software for Education, 1(33). https://doi.org/10.12753/2066-026x-15-020

Kaimara, P., Oikonomou, A., & Deliyannis, I. (2022). Could virtual reality applications pose real risks to children and adolescents? A systematic review of ethical issues and concerns. Virtual Reality, 26, 697–735. https://doi.org/10.1007/s10055-021-00563-w

Loughlin, E., & Mentzer, K. (2026). Adapting CAMIL for K–12 world language learning: A qualitative study of immersive VR in middle school Spanish. Information Systems Education Journal, 25(1), 55–65. https://doi.org/10.62273/DIWX3099

Ma, W., Adesope, O. O., Nesbit, J.C., &Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of Educational Psychology, 106(4), 901-918. http://dx.doi.org/10.1037/a0037123

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., Makransky, G., & Parong, J. (2023). The promise and pitfalls of learning in immersive virtual reality. International Journal of Human–Computer Interaction, 39(11), 2229–2238. https://doi.org/10.1080/10447318.2022.2108563

Mentzer, K., Frydenberg, M., & Li, S. (2025). Exploring VR-enhanced learning in business education: A multi-site study. Information Systems Education Journal, 23(5), 30–43. https://doi.org/10.62273/WOFR1838

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. Consciousness and Cognition, 21(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Patterson, A., Temple, C., Anderson, N., Rogalski, C., & Mentzer, K. (2025). The Virtual Stage: Virtual reality integration in effective speaking courses. Information Systems Education Journal, 23(4), 57–68. https://doi.org/10.62273/IMLU5983

ISSN: 2473-4901
v12 n6706

Appendix A.

Full Pooled Repeated-Attempt Results

Metric	n	Attempt 1	Attempt 2	Mean change	Direction	p-value
Overall score	208	6.43	6.71	0.283	Improved	0.000*
Content score	208	6.38	6.68	0.301	Improved	0.000*
Delivery score	208	6.85	6.99	0.143	Improved	0.002*
Eye contact (%)	203	0.8	0.85	0.053	Improved	0.000*
Body language (%)	203	0.13	0.15	0.019	Improved	0.039*
Listenability (%)	203	0.77	0.75	-0.026	Degraded	0.000*
Words per minute	203	129.20	148.5	19.3	Higher	0.000*
Filler words	203	7.58	6.4	-1.18	Improved	0.042*
Duration (sec)	208	406.1	374.8	-31.3	Improved	0.001*

ISSN: 2473-4901
v12 n6706

Appendix B.

Full Activity-Level Repeated-Attempt Results

Analysis group	Metric	n	Attempt 1	Attempt 2	Mean change	Direction	p-value
1st Year - Activity 1	Overall score	49	6.57	6.82	0.25	Improved	0.102
1st Year - Activity 1	Content score	49	6.57	6.8	0.25	Improved	0.125
1st Year - Activity 1	Delivery score	49	6.87	7.02	0.15	Improved	0.044*
1st Year - Activity 1	Eye contact (%)	49	0.78	0.86	0.07	Improved	0.007*
1st Year - Activity 1	Listenability (%)	49	75.2%	72.9%	-2.3%	Lower	0.038*
1st Year - Activity 1	Words per minute	49	126	150	24	Faster	0.000*
1st Year – Activity 1	Filler words	49	6	5.08	-0.92	Fewer	0.24
1st Year - Activity 1	Duration (sec)	49	381	332	-49	Shorter	0.002*
1st Year - Activity 2	Overall score	75	6.59	6.94	0.36	Improved	0.001*
1st Year - Activity 2	Content score	75	6.55	6.94	0.4	Improved	0.001*
1st Year - Activity 2	Delivery score	75	6.95	7.02	0.07	Improved	0.356
1st Year - Activity 2	Eye contact (%)	75	79.1%	83.2%	0.04	Improved	0.029*
1st Year - Activity 2	Listenability (%)	75	81.9%	77.8%	-4%	Lower	0.000*
1st Year - Activity 2	Words per minute	75	127.3	147.2	19.9	Faster	0.000*
1st Year - Activity 2	Filler words	75	4.65	4.59	-0.07	Fewer	0.909
1st Year - Activity 2	Duration (sec)	75	337.1	297.3	-39.8	Shorter	0.005*
3rd Year - Interview	Overall score	46	6.25	6.49	0.24	Improved	0.083
3rd Year - Interview	Content score	46	6.17	6.44	0.27	Improved	0.076
3rd Year - Interview	Delivery score	46	6.88	6.9	0.03	Improved	0.7274
3rd Year - Interview	Eye contact (%)	41	90.5%	92.9%	2.4%	Improved	0.5108
3rd Year - Interview	Listenability (%)	41	66.7%	67.2%	0.5%	Higher	0.762
3rd Year - Interview	Words per minute	41	153.3	154.1	0.8	Faster	0.912
3rd Year - Interview	Filler words	41	18.83	14.34	-4.49	Fewer	0.07
3rd Year - Interview	Duration (sec)	46	644.7	647.9	3.2	Longer	0.912

ISSN: 2473-4901
v12 n6706

Appendix C.

Qualitative Coding Framework Summary

Qualitative data were analyzed using a hybrid deductive-inductive thematic approach. The research team began with a predefined set of codes derived from the study's research questions, prior immersive learning literature, and an initial review of the student reflections. Coders were allowed to assign more than one code to a response segment when appropriate.

- **Perceived Value / Usefulness:** Comments describing the VR activity as helpful, valuable, beneficial, or worthwhile.
- **Realism / Authenticity:** Comments addressing whether the simulation felt realistic, immersive, artificial, robotic, or unlike a real interaction.
- **Confidence / Anxiety:** Comments reflecting increased confidence, reduced nervousness, greater comfort, or lower anxiety during or after the activity.
- **Feedback / Coaching Value:** Comments emphasizing the usefulness of the AI coach, feedback session, scoring, or reflective guidance after the roleplay.
- **Perceived Improvement / Repetition:** Comments indicating growth across attempts, improvement from one session to the next, or application of prior feedback.
- **Communication Self-Awareness:** Comments identifying communication strengths or weaknesses such as pacing, eye contact, tone, filler words, listening, or clarity.
- **Transfer to Real-World Situations:** Comments linking the experience to future use in teamwork, difficult conversations, interviews, leadership, class projects, or workplace communication.
- **Technical / Design Friction:** Comments describing barriers related to the technology or activity design, including repetitive questioning, awkward AI interaction, discomfort, or difficulty using the system.
- **Preference for Human Interaction:** Comments suggesting that real people, live mock interviews, or in-person conversations would be more valuable than AI-based interaction.