

An Early-Stage Feasibility and Usability Pilot for an Immersive Serious Game to Support English Spelling and Semantic Skills Using AI

Benjamin Pedley and Marco Gilardi

Faculty of Computing, Sciences and Engineering, Department of Computing and Digital Technologies,
University of the West of Scotland, Paisley, UK

bennypedley78@gmail.com

Marco.Gilardi@uws.ac.uk

Abstract: In recent years, the rise of Artificial Intelligence (AI) has created new opportunities to build tools that support adult language acquisition and literacy skills. At the same time, Immersive Serious Games (iSG) have been shown to be more effective in increasing student engagement and motivation than traditional classroom methods. By combining iSG and AI, accessible spaces where students can practice their language skills, improve their spelling, and grow their semantic understanding can be developed. To test this hypothesis, a virtual reality (VR) iSG combined with an artificial intelligence (AI) disembodied agent was created to support the learning of English language skills. This AI-VR iSG allows students to practice at their own pace through mini games. For instance, a “Word Cloud” game makes the act of spelling more physical and tangible, while a “Basketball” game makes semantic understanding more engaging. In addition to the mini games, an AI agent act as a virtual tutor that is always available to the students to help with spelling tasks, reducing anxiety for students that may feel anxious during face-to-face sessions with instructors. A pre-test post-test experimental design was used to measure spelling skills and semantic understanding improvement. A spelling test was administered asking participants to spell 32 words ranging from short 3-letter words to more complex 10-letter words measuring improvements of words spelling before and after the use of the iSG. Finally, a questionnaire measuring accessibility and engagement of the serious game was also administered in the post iSG session. Results indicated medium effect size improvements in spelling proficiency and VR engagement, alongside noted challenges regarding AI integration responsiveness. Ultimately, this paper shows that VR and AI are promising technologies to provide a supportive, asynchronous space for learning that helps students gain confidence while also making the workload easier to support teachers.

Keywords: Artificial intelligence, Virtual reality, Immersive serious games, Spelling, Language skills

1. Introduction

Proficiency in language skills is fundamental to ensure societal functioning; however, the rate of literacy worldwide is at 87.74% (World Population Review, 2025), with levels of literacy varying drastically between the global south and the global north. Even within the UK there are disparities in literacy levels between the four countries (England 82%, Scotland 73.3%, Wales 88%, Northern Ireland 83%) (National Literacy Trust, 2024) indicating a critical need for innovative educational interventions to teach literacy skills.

Immersive Serious Games (iSG) have been shown to be more effective in fostering higher positive affect compared to traditional teaching (Allcoat and von Mühlenen, 2018, Lamb et al., 2018). By combining iSG and AI, accessible spaces where students can practice their language skills, improve their spelling, and grow their semantic understanding can be developed.

This study discusses a Virtual Reality (VR) and Artificial Intelligence (AI) assisted Serious Game (SG) for teaching spelling and semantics in the English language within a controlled experimental setting. The proposed system integrates an AI agent within an immersive virtual environment, facilitating minigames focused on spelling, grammar, and semantics. This early-stage pilot evaluates the tool for a generalized population of "adult language learners," which intentionally encompasses both native speakers seeking literacy improvement and non-native English speakers.

2. Related Work

VR has been shown to be effective in the teaching of general English (Pan et al. 2021) and written English as a foreign language (Feng and Ng, 2023), and has the potential to significantly enhance student engagement, enjoyment, and motivation compared to traditional methods (Chen, 2022; Parong and Mayer, 2018).

The availability of Artificial Intelligence (AI) has created new opportunities to build tools that support literacy and students through personalized feedback. AI can provide support mechanisms that improve learner experience (Lamnai and Elmhouti, 2025), offering pedagogical potential through personalized feedback, enhancing motivation and reflective thinking (Wang et al., 2024; Chalouy and Tsai, 2025, Yu, 2026). However,

technical limitations often hinder effective language acquisition, with latency and verbosity being key friction points (Chalouy and Tsai, 2025). Furthermore, AI can introduce potential bias and misinformation generation (Siau and Wang, 2020; Yao et al., 2024).

Technology enhanced language education has been investigated by integrating VR and AI and game-based learning (GBL). Ward, Xu, and Dhonnchadha (2025) propose an VR-AI serious game system for teaching the Irish and other Celtic languages, showing positive reception within schools, despite the limited content used for testing. A limitation of Ward, Xu, and Dhonnchadha (2025) was the ability of real time speech interaction, this has been explored by Divekar et al. (2022) in teaching Chinese as a foreign language showing that this modality provides statistically significant results for improved vocabulary, comprehension and conversational skills, despite the small sample size. Lou (2025) shows that VR-AI integration yields positive results in language acquisition. These results are supported by Bazargani et al. (2025) gamified VR-AI environment in the context of spelling for English colour word acquisition by young learners. Bazargani et al. (2025) found that the gamified VR application improved spelling performance and engagement. However, Bazargani et al. (2025) study was limited to ten colour words. A more general approach was taken by Li (2021), which introduced AI-VR environments for the teaching of English at college level, showing learning improvements compared to the control test.

While previous research has established the individual benefits of VR and AI, and preliminary studies like Bazargani et al. (2025) have explored this intersection for children, a clear gap remains regarding open-ended, AI-assisted VR spelling and semantics for adult learners.

Because literature suggests VR enhances engagement and embodied learning (Parong and Mayer, 2018), we designed mechanics that require physical interaction (manipulating 3D letters, throwing basketballs) to support spelling and semantic skills, rather than relying on traditional 2D interfaces.

3. Methodology

This study adopts a mixed-methods research design, integrating quantitative experimental data with qualitative survey responses to evaluate the AI-assisted Virtual Reality (VR) serious game. Ethical approval was obtained from the UWS Ethics Board, and the experiment followed the University ethical guidelines.

Recruitment targeted university-wide channels, yielding 16 participants with a mean age of 23.66 years. The cohort included six non-native English speakers and one participant reporting a learning difficulty. The experimental procedure commenced with a spelling test with 8 levels of increasing difficulty, consisting of 32 words, with 4 words being used for each word length starting from 3 letters to 10. The sessions were recorded to assess completion time and error rates. Participants subsequently engaged with the VR-AI game environment for approximately thirty minutes before undertaking a second, distinct spelling assessment using novel words.

Following Bloomer (1956) and Spencer (2002), who identified word length as a significant predictor of orthographic complexity, word length was used to establish a baseline framework and construct test items, despite limitations regarding phonemic metrics.

A survey featuring five points Likert-scale items regarding accessibility, engagement, and AI utility, alongside open-ended questions on overall recommendations was used to capture subjective perceptions relative to their prior learning experiences. Open-ended qualitative data was used to explore, following Brown and Clarke (2006) perceptions of AI utility and immersion. Descriptive statistics, normality testing via Shapiro-Wilk tests, Q-Q plots, and paired t-tests were conducted between pre- and post-test interventions. Cohen's *d* coefficients were calculated to determine effect sizes. The analysis was done by the primary author of the study and reviewed by the secondary author.

4. Game Design and Implementation

The development of the immersive serious game followed an Agile iterative (Amoros and Krasamo, 2025), ensuring flexibility in refining user experience elements once initial prototypes were constructed, allowing for adjustments based on technical feasibility and usability testing.

Following Ensmann and Mullins (2026) and Chen et al. (2018), which showed increases in learning motivation and overall enhancement of the learning experience over desktop gaming, the design decision was made to design the game as immersive VR.

The conceptual design aimed to create an engaging environment suitable for adult learners. To enhance immersion, the game employed high-fidelity 3D graphics within a first-person perspective. The aesthetic prioritised bright visuals and clarity to minimise distraction during learning tasks. The gameplay structure consisted of four distinct levels: a tutorial, a hub for navigation, a spelling minigame, and a semantics game. Following the completion of the tutorial, participants accessed a central hub area which served as a navigation node, allowing access to minigame via pathways leading to specific thematic environments Figure 1.

Technical implementation was executed using the Unity Engine 6 streamed on Meta Quest 2 headsets supported by an AMD Ryzen 3600 CPU and AMD Radeon 6600XT GPU. To mitigate potential motion sickness, a vignette effect was applied during movement.

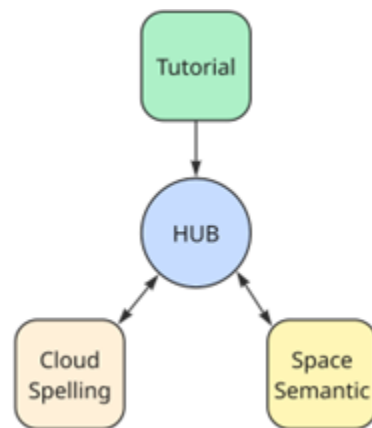


Figure 1: Virtual Environment Structure

Artificial Intelligence (AI) was integrated to provide scaffolding. The Google Gemini 2.5 Flash API was selected over local models due to superior processing capabilities and reliability for this context. The AI integration architecture data flow involved the game client sending prompts via the API, receiving text responses, and converting them into audio using Text-to-Speech (TTS) APIs before playback, the API default parameters were left unchanged (temperature = 1, topP = 0.95, topK = 64). A custom script managed this communication using several role-playing prompts such as *“You are a helpful, patient, language and spelling assistant inside a VR game that uses standard Unity VR controller...”* and *“The player will be given a word to spell using the letters they can see in front of them. Let the player know that if they ever get stuck or need any help they can ask and you will be happy to assist them.”* This was paired with smaller prompts dealing with questions and words for the minigames. To ensure AI safety, Gemini content filtering configurations was used. Gemini 2.5 Flash had a reasoning accuracy of 82.8% (GPQA (diamond)) and a factuality accuracy of 85.3% (FACTS Grounding) (Comanici et al., 2025), moreover, the Gemini 2.5 Flash API was free for students.

At the start of the game the user sends data, such as their names and words they would like help with spelling while playing, which the AI remembers throughout the playthrough. Upon restart of the game data the chat history is wiped. Windows Speech Recognition was integrated via a secondary script to enable voice commands when the player activated the X button, allowing for hands-free interaction with the AI assistant.

Environmental design and level logic were implemented to support specific learning objectives. The Tutorial Level (Figure 2, top left), was designed to acclimatize players to physics and controls through guided AI interaction. The Hub level functioned as a central navigation point (Figure 2, top right). Two primary gameplay loops were developed. The 'Cloud' level focused on spelling mechanics; players retrieved floating letters to construct target words, requiring participants to physically reach for and manipulate each letter in 3D space, this is intended to reinforce spelling learning through motor and visual engagement (Figure 2, bottom left). The 'Space' level consisted of a basketball quiz game regarding grammar and semantics, visual fidelity remained consistent with a space theme (Figure 2, bottom right) and allowed for participants to get a physically interactive semantic experience to reinforce learning through active responses.

User Interface (UI) accessibility was prioritized through simple trigger-based menu interactions. To ensure continuity of immersion and reduced cognitive load upon game entry, the Start Menu was moved from the space environment to the Tutorial level setting, preventing sudden environmental disorientation (Figure 2, top right). Audio design utilized both diegetic soundscapes, such as office ambience for the tutorial and wind sounds for the cloud level, alongside non-diegetic background tracks in the space arena to maintain atmospheric cohesion.

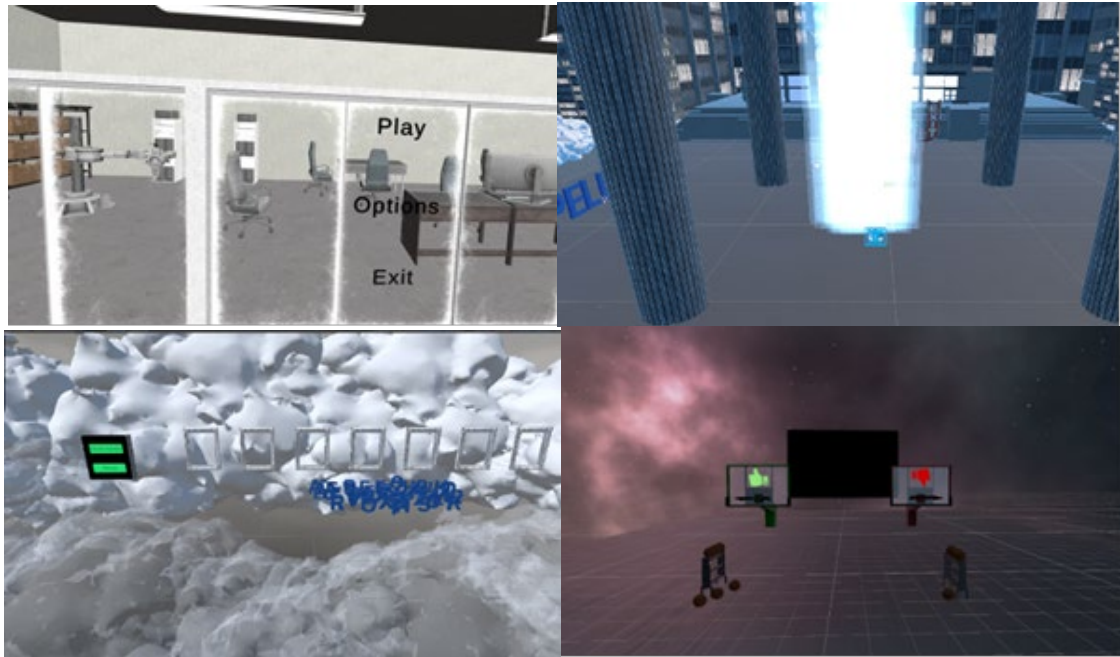


Figure 2: Top Left: Tutorial Environment; Top Right: Hub Environment; Borrom Left: Cloud Spelling Mini-game; Bottom-Right: Space Semantic Mini-Game

5. Results

The results presented in this section detail the empirical findings derived from evaluating the AI-assisted Immersive Serious Game. The data analysis comprises quantitative metrics regarding spelling proficiency and task completion time, alongside survey responses, quantifying user experience and self-perceived improvement.

5.1 Quantitative Data Analysis

The experimental phase involved 16 recruited participants; however, data quality protocols required the exclusion of two participants to maintain statistical integrity. Prior to the primary analysis, the data were screened for univariate outliers using the Interquartile Range (IQR) method, defined as values falling 1.5 times the IQR below the first quartile or above the third quartile. One participant was identified as an outlier regarding their Time Difference (Difference = 65s), as this value significantly exceeded the upper boundary of 44.22s. No outliers were detected for the Mistakes Difference (normal range: -5.75 to 4.25). This single outlier was excluded from subsequent analyses. A second participant was excluded due to inaccuracies in their recording process during the session. Consequently, the final analysis relied on data from 14 participants. The experiment measured two primary dependent variables: the frequency of spelling errors and the duration required to complete the assessment.

Spelling mistakes rates: Results indicated a mean reduction from the pre-test ($M = 3.86$) to the post-test ($M = 3.07$), yielding a mean difference of -0.79 . Caution should be applied in interpreting the results as the difficulty equivalence between the pre- and post-test word lists was not established; see Discussion. To evaluate the distribution of these differences, a Shapiro-Wilk test was administered ($W = 0.898$, $p = 0.104$). As the p -value exceeded 0.05, the assumption of normality was upheld, as visually confirmed by the Q-Q plot and bar-chart analysis (Figure 3). A paired-samples t -test was then conducted to determine statistical significance with an alpha level of 0.05. The test yielded a t -statistic of 1.924 ($df = 13$) and a p -value of 0.077, failing to reject H_0 at the 0.05 level. However, the confidence interval of $[-1.67, 0.10]$ suggests a trend toward improvement, with zero approaching significance. This was further contextualized by Cohen's d coefficient of 0.514 (95% CI: $[-0.05, 1.06]$), which indicates a medium effect size, suggesting a trend toward improvement that warrants further investigation.

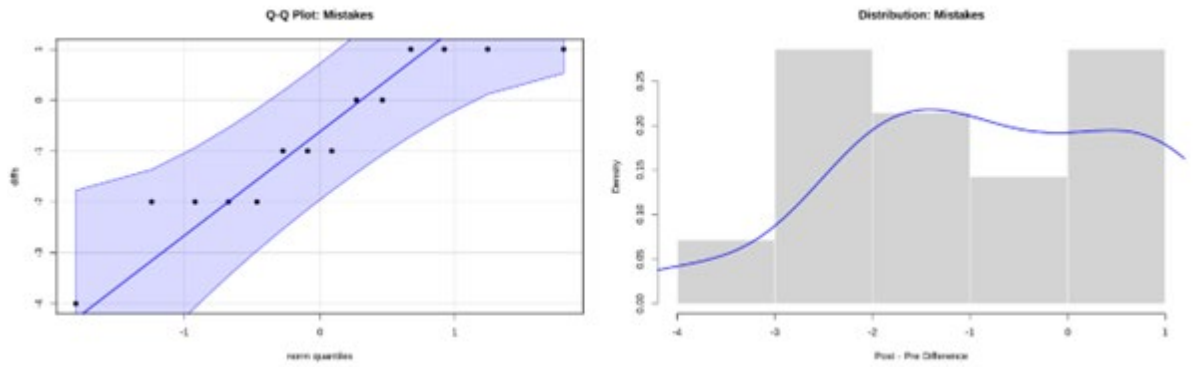


Figure 3: Left: Q-Q plot of the distribution of the pre-post-test difference for the ‘mistakes’ independent variable; Right: Bar chart of the pre-post-test difference for the ‘mistakes’ independent variable

Completion time (measured in seconds): Pre-test times averaged 194.01s (SD = 36.65), decreasing to 185.63s (SD = 34.33) post-intervention, resulting in a mean difference of -8.38s. Normality was confirmed via a Shapiro Test ($W = 0.944$, $p = 0.472$) and Q-Q plot and bar chart analysis Figure 4. A paired t-test returned a t-value of 1.673 ($df = 13$) with a p-value of 0.118, failing to reject the null hypothesis at alpha 0.05. The confidence interval for this metric was [-19.20, 2.44], crossing zero. Nevertheless, Cohen’s d was calculated at 0.447 (95% CI: [-0.11, 0.99]), a medium effect size regarding efficiency gains. These results imply that while time savings were not statistically significant, the intervention may reduce task duration on a practical level.

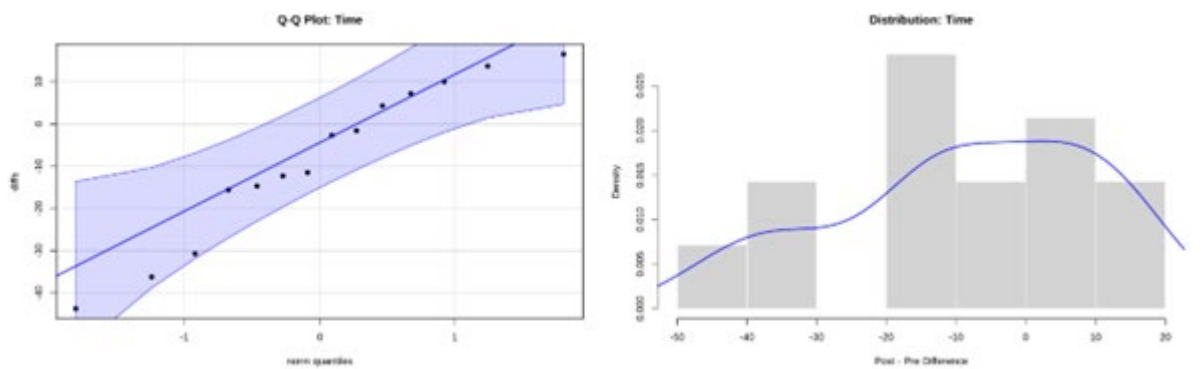


Figure 4: Left: Q-Q plot of the distribution of the pre-post-test difference for the ‘time’ independent variable; Right: Bar chart of the pre-post-test difference for the ‘time’ independent variable

5.2 Survey Data Analysis

All 16 participants completed the post-game survey. Demographic data revealed a mean age of 23 years, with 6 non-native English speakers and one participant reporting learning difficulties. Participants rated the game’s attributes using a 5-point Likert scale. Accessibility was highly rated by 12 participants (Very Accessible), while intuitiveness received strong endorsement from 13 participants (Very Intuitive). Engagement was predominantly positive, with 15 participants finding the experience fun or very fun. When comparing the VR game against traditional methods previously employed by participants (e.g., phonetics, mnemonics), there were no significant declines in ratings for intuitiveness or accessibility; however, engagement ratings showed a slight reduction, indicating that while the tool was accessible, it may require additional onboarding compared to prior familiar methods.

Perceived self-efficacy outcomes were notably positive. Twelve participants reported feeling their spelling ability improved (9 small increase, 3 large increase). Regarding semantic understanding, 14 participants felt their skills were affected positively (10 small increase, 4 large increase), followed by grammar comprehension, where 10 out of 15 respondents felt their understanding had improved. The immersive nature of the environment was highly valued, with 15 participants affirming that VR enhanced their learning experience. Ratings for improvement of quality within the VR context averaged high (Mean 4 out of 5). However, the AI assistant received mixed utility ratings ($M = 3.20$, $SD = 1.21$): 6 rated it as having very limited helpfulness, 3 rated it neutrally, 3 found it helpful, and 3 rated it as very helpful, indicating variability in user interaction with the agent.

5.3 Qualitative Data Analysis

Analysis of the open-ended survey responses revealed distinct patterns regarding the VR environment, the AI agent, and general feedback. Three primary themes emerged concerning the VR interface. First, immersion was viewed as a positive driver for engagement; most participants explicitly contrasted the game with traditional learning methods, noting that VR was more entertaining (e.g., "I think it's extremely interactive as compared to non-VR experiences due to its high intuitive it really makes you connected to the experience" (p1)). Second, the removal of environmental distractions was frequently cited as a facilitator for focus. Conversely, some participants reported experiencing disorientation (e.g., "I think a stationary monitor with a keyboard and mouse is less disorientating to use than VR" (p7)), which detracted from their ability to learn.

Five themes characterized the AI assistant's interaction. On the positive side, the agent provided clear instructions and effective definitions for difficult words, acting as a useful guide. However, significant friction points were identified. Participants noted that the AI spoke excessively, interrupting cognitive processing time with lengthy explanations. For example, one participant noted, "he was good for defining words and for simple assistance but he's rather repetitive" (p8), while another stated, "the AI voice was very annoying. Wouldn't stop talking" (p4). Furthermore, technical latency was a barrier; users reported slow response times where they had to wait for feedback. Additionally, some participants found the agent non-functional during gameplay, as it failed to process voice input, limiting real-time interaction opportunities.

Finally, while the core game mechanics were described as fun and engaging, there was a consensus that the AI assistant should become faster and more consistent in its responses to ensure a seamless learning flow without interruption. These qualitative findings underscore the necessity of refining agent responsiveness and dialogue management to maximize the educational benefits identified by participants.

6. Discussion

The analysis reveals promising trends regarding spelling acquisition, user accessibility, and semantic understanding, while also highlighting methodological constraints and technical limitations that inform future iterations of the game.

Regarding the first research question concerning the impact on spelling abilities, the results indicate a positive trend despite statistical non-significance. Survey data showed that 75% of participants perceived an increase in their spelling skills post-intervention. This was corroborated by experimental metrics; quantitative analysis showed a mean reduction in both errors and completion time between pre- and post-test assessments. However, the study was underpowered, precluding statistical significance, but this did not negate the efficacy observed in similar contexts where engagement drove performance (Allcoat and von Mühlennen, 2018). Moreover, Cohen's *d* coefficient indicated a medium effect size (approx. 0.5), a promising preliminary signal that the intervention possessed some practical value even though the *p*-values exceeded the conventional alpha level of 0.05 ($p = .077$ for mistakes; $p = .118$ for time). Limitations existed regarding the assessment framework; utilizing word length as the sole predictor of difficulty was deemed rudimentary by Spencer (2002). Because the strict equivalence of the pre- and post-test word sets was not established, it is possible that the observed reduction in spelling errors reflects inequalities in test difficulty rather than genuine language acquisition. Future iterations should incorporate phonemic complexity or frequency counts to ensure accuracy in measuring difficulty independent of word length.

The second research question investigated accessibility and usability. A contrast between VR immersion and AI assistance emerged. Quantitative survey data revealed that 93.8% of participants felt VR improved the learning experience, primarily due to the removal of external distractions and user perceived increased engagement compared to traditional desktop methods. This supports Checa et al. (2021) and suggests that VR environments are inherently more motivating (Parong and Mayer, 2018). Specifically, the requirement to physically grasp 3D letters and throw basketballs aligns with principles of embodied cognition, where motor-kinaesthetic engagement actively reinforces memory encoding and semantic retrieval, offering mechanical advantages over passive 2D interfaces. However, Likert-scale comparisons between game accessibility and traditional methods yielded near-neutral results. While qualitative data suggested users felt the game was accessible due to intuitive controls, quantitative results warrant further investigation, potentially stemming from survey timing or bias. The AI assistant component presented mixed outcomes. While some participants found the feedback clear and useful (Wang et al., 2024), the majority rated the agent as not useful (3/5 or less). This friction was due to response times latency and verbosity caused by poor prompt planning. This aligns with Chalouy and Tsai (2025),

highlighting that while AI feedback is beneficial, implementation requires faster inference and concise dialogue to maintain flow.

The third research question examined semantic understanding. Due to time constraints, a controlled experiment measuring semantic acquisition was not conducted; therefore, data relied on self-reporting. Nevertheless, 87.5% of participants felt their understanding of word meanings had improved, with over 60% noting significant gains. A smaller subset reported similar gains in grammar knowledge. The consistency of these self-reported improvements mirrors the objective spelling gains (medium effect size), suggesting that while experimental verification is required, the mechanism of learning within the game likely extends across language domains. However, without a pre-post-test specifically for semantics, these claims remain correlative rather than causal.

In conclusion, this study indicates potential for AI-assisted VR games as viable tools for adult language acquisition. While engagement and immersion were leveraged to improve spelling accuracy, technical refinements are required regarding AI responsiveness and the sophistication of difficulty metrics. These insights provide a clear roadmap for developing more robust educational interventions that balance technological innovation with pedagogical efficacy.

7. Conclusions and Limitations

This study investigated a Virtual Reality (VR) and Artificial Intelligence (AI)-assisted serious game for enhancing adult language skills, addressing three research questions regarding spelling accuracy, accessibility, and semantic understanding. Findings provide initial evidence regarding spelling proficiency, though methodological limitations necessitate further investigation. Quantitative analysis indicated a medium effect size in improving spelling accuracy and speed; however, improvements did not reach statistical significance due to the reduced sample size. These preliminary results provide early-stage encouragement for the system's potential, pending verification in larger, controlled studies. Regarding accessibility, results were mixed. While participants reported that VR integration enhanced the learning experience by removing external distractions, the AI assistant hindered usability due to slow response times and verbosity. Conversely, self-reported gains in semantic understanding were notably high, with many participants perceiving improvements across multiple language domains, although these claims remain unverified without specific experimental testing.

Since this is an early stage limited time pilot word length was the only difficulty criterion used, 4 words were used at each difficulty level to increase consistency between tests. However, more tests should be conducted with more sophisticated methods using more phonemic methods.

Given that this study is an early-stage pilot, we were unable to test individual aspects such as the AI, VR, or specific game mechanics, nor can we rule out the effects of repeated testing or simple practice. Because there was no control condition, strong conclusions regarding comparative educational effectiveness cannot be drawn from this data. The observed change in the results has been preliminarily accredited by a combination of AI, VR, and game mechanics. However, further testing is needed on individual aspects to see to what extent they influence results.

Based on these outcomes, the following recommendations are proposed to address this study limitations. Increasing the sample size is critical to enhance statistical power regarding spelling efficacy in adult populations. Technical refinements are essential regarding AI integration, optimizing model inference speeds to reduce latency and refine prompt engineering strategies to prevent cognitive flow interruptions. Furthermore, incorporating phonemic or frequency data will improve the validity of future spelling accuracy assessments. Reduce biasing of pre and post-tests. Increase focus on semantic understanding is recommended to enhance empirical validation of linguistic acquisition. Finally, implement a control group to increase fairness and accuracy of comparisons. Applying these steps would transform the current prototype into a scalable educational tool that effectively leverages VR immersion and AI support without compromising usability or instructional integrity.

Ethics Declaration: Ethical approval for this study was obtained from the Computing, Engineering and Physical Sciences SAIEC at the University of the West of Scotland, where the user evaluation took place. Ethics Application number: 24172 Approval date: 28/11/2025

AI Declaration: Due to the first author's dyslexia, a locally run AI (Qwen 3.5-9b) was used to support writing. Sentences originally written by the first author were fed to the AI with the request to polish them to an academic standard. The sentences were then re-read by the first and second author and adapted to the paper.

References

- Alfadil M. (2020) "Effectiveness of virtual reality game in foreign language vocabulary acquisition", *Computers & Education*, Volume 153, 03893, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2020.103893>
- Allcoat, D. and von Mühlenen, A. (2018) "Learning in virtual reality: Effects on performance, emotion and engagement" *Research in Learning Technology*, 26(0) <https://doi.org/10.25304/rlt.v26.2140>
- Amoros, J. and Krasamo (2025) "Agile Development". Online . Available at: <https://www.krasamo.com/agile-development-process> [Accessed 7 Apr. 2026]
- Bazargani J. S., Sadeghi-Niaraki A., Shi X. and Choi S. -M. , "Adaptive Spelling in Immersive Reality: The Impact of Gamified VR and LLMs on Young Learners' English Color Word Acquisition", in *IEEE Access*, vol. 13, pp. 131789-131802, 2025, <https://doi.org/10.1109/ACCESS.2025.3591114>
- Bloomer, R.H. (1956) "Word Length and Complexity Variables in Spelling Difficulty." *The Journal of Educational Research*, [online] 49(7), pp.531–535. <https://doi.org/10.2307/27529865>
- Braun, V. and Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2), pp.77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Chalouy, N. and Tsai, A.-C. (2025) "Enhancing Language Learning Through AI-Driven Feedback and Collaborative Virtual Reality". In: *Lecture Notes in Computer Science*. 8th International Conference, ICITL 2025 Oslo, Norway, pp.299–308. https://doi.org/10.1007/978-3-031-98197-5_32
- Checa, D., Miguel-Alonso, I. and Bustillo, A. (2021) "Immersive virtual-reality computer-assembly serious game to enhance autonomous learning". *Virtual Reality*, pp.1–18. <https://doi.org/10.1007/s10055-021-00607-1>
- Chen, Z. (2022) "Exploring the application scenarios and issues facing Metaverse technology in education, Interactive Learning Environments" *Interactive Learning Environments*, 32(5), pp.1975–1987. <https://doi.org/10.1080/10494820.2022.2133148>
- Chen, X., Chen, Z., Li, Y., He, T., Hou, J., Liu, S. and He, Y. (2018) "ImmerTai: Immersive motion learning in VR environments". *Journal of Visual Communication and Image Representation*, 58(58), pp.416–427. <https://doi.org/10.1016/j.jvcir.2018.11.039>
- Comanici, G., Bieber, E., Schaeckermann, M., Pasupat, I., Sachdeva, N., Dhillon, I. and Ram, O., et al. (2025) "Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities". *arXiv preprint*, arXiv:2507.06261. <https://arxiv.org/abs/2507.06261>
- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., ... Braasch, J. (2022) "Foreign language acquisition via artificial intelligence and extended reality: design and evaluation", *Computer Assisted Language Learning*, 35(9), 2332–2360. <https://doi.org/10.1080/09588221.2021.1879162>
- Ensmann, S. and Mullins, A. (2026) "Motivation and learning in serious games across virtual reality and desktop environments". *International Journal of Game-Based Learning*, 16(1) <https://doi.org/10.4018/IJGBL.415263>
- Feng B and Ng LL (2023) "Facilitating writing performance of EFL learners via virtual reality: Immersion, presence, embodiment". *Front. Psychol.* 14:1134242. <https://doi.org/10.3389/fpsyg.2023.1134242>
- Lamb, R.L., Annetta, L., Firestone, J. and Etopio, E. (2018) "A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations" *Computers in Human Behavior*, 80, pp.158–167. <https://doi.org/10.1016/j.chb.2017.10.040>
- Lamnai, A. and Elmhouti, A. (2025) "Artificial Intelligence in Educational Serious Games". *Advances in Educational Technologies and Instructional Design*, pp.231–264. <https://doi.org/10.4018/979-8-3693-8965-2.ch012>
- Li M. (2021) "An Immersive Context Teaching Method for College English Based on Artificial Intelligence and Machine Learning in Virtual Reality Technology", *Mobile Information Systems*, 2021, 2637439 <https://doi.org/10.1155/2021/2637439>
- Lou Y. (2025) "The impact of virtual reality environments on English language acquisition: Innovative immersive learning technologies for communication skills development." *Journal of Computational Methods in Sciences and Engineering*. 25(5):4616-4632. <https://doi.org/10.1177/14727978251337950>
- National Literacy Trust (2024) "Adult Literacy, National Literacy Trust". National Literacy Trust. Available at: <https://literacytrust.org.uk/parents-and-families/adult-literacy/>.
- Pan Z., Sun Y., Yao Z. W. and Li M. (2021)"Application of Virtual Reality in English Teaching," 2021 3rd World Symposium on Artificial Intelligence (WSAI), Guangzhou, China <https://doi.org/10.1109/WSAI51899.2021.9486322>
- Parong, J. and Mayer, R.E. (2018) "Learning science in immersive virtual reality". *Journal of Educational Psychology*, 110(6), pp.785–797. <https://doi.org/10.1037/edu0000241>
- Siau, K. and Wang, W. (2020) "Artificial Intelligence (AI) Ethics". *Journal of Database Management*, [online] 31(2), pp.74–87. <https://doi.org/10.4018/jdm.2020040105>
- Spencer, K. (2002) "English spelling and its contribution to illiteracy: word difficulty for common English words". *Reading*, 36(1), pp.16–25. <https://doi.org/10.1111/1467-9345.00180>
- Wang, W.-S., Lin, C.-J., Lee, H.-Y., Huang, Y.-M. and Wu, T.-T. (2024) "Integrating feedback mechanisms and ChatGPT for VR-based experiential learning: impacts on reflective thinking and AIoT physical hands-on tasks". *Interactive Learning Environments*, 33(2), pp.1770–1787. <https://doi.org/10.1080/10494820.2024.2375644>
- Ward M, Xu L., and Dhonnchadha E. U. (2025) "A Pragmatic Approach to Using Artificial Intelligence and Virtual Reality in Digital Game-Based Language Learning". In *Proceedings of the 5th Celtic Language Technology Workshop*, pages 27–34, Abu Dhabi [Virtual Workshop]. International Committee on Computational Linguistics.

- World Population Review (2025) "Literacy Rate by Country", worldpopulationreview.com. Available at:
<https://worldpopulationreview.com/country-rankings/literacy-rate-by-country>
- Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z. and Zhang, Y. (2024) "A survey on Large Language Model (LLM) security and privacy: The Good, The Bad, and The Ugly". High-Confidence Computing, 4(2), p.100211.
<https://doi.org/10.1016/j.hcc.2024.100211>
- Yu, M. (2026) "Optimizing EFL vocabulary acquisition: a randomized controlled mixed-methods investigation of artificial intelligence-driven incidental, contextual, and multimodal strategies". Educ Inf Technol 31, 53–97.
<https://doi.org/10.1007/s10639-025-13803-2>