Immersive Learning Research Network

LexiLearn: Adaptive AI-Driven Virtual Reality for Multisensory Literacy Instruction and Pedagogical Support

Lucas Yeykelis, Marie P.P.M. Holm, Neon Cao, Nasir Grant, Christopher Eagar, Bryson Rudolph, Anna C.M. Queiroz
University of Miami, Coral Gables, FL, USA
aqueiroz@miami.edu

Abstract. We present LexiLearn, an AI-driven, immersive virtual reality intelligent tutoring system designed for foundational literacy acquisition. The system integrates real-time learner performance analytics, adaptive content sequencing, and multimodal interaction to provide personalized spelling and phonics practice within an embodied virtual environment. An AI engine continuously models learner proficiency by tracking accuracy, error patterns, and response timing, dynamically adjusting task difficulty, word selection, and repetition schedules to maintain an optimal challenge level. Generative AI produces categorized relevant visual imagery to reinforce word-meaning associations, while phoneme–grapheme interactions are synchronized with multisensory haptic feedback delivered via haptic gloves. An AI-enhanced teacher dashboard translates learner model outputs into interpretable pedagogical insights through conversational AI, including detecting misconception patterns and recommending classroom interventions. This closes the loop between automated adaptation and human instructional decision-making, positioning the system as a hybrid intelligent tutoring and teacher-support platform.

Keywords: Immersive Learning, Foundational Literacy, Embodied Interaction, Intelligent Tutoring Systems, Virtual Reality, Generative AI.

1 Introduction

Foundational literacy remains one of the most pressing challenges in global education. Despite sustained investment in reading instruction, millions of children continue to struggle with basic decoding and spelling skills, limiting their academic trajectories and opportunities for social participation [14]. Traditional instructional approaches often rely predominantly on visual and auditory modalities and struggle to deliver individualized feedback at scale [7]. LexiLearn targets learners spanning elementary through middle school, including older students who struggle with foundational decoding and spelling skills.

Research in embodied cognition and multisensory learning indicates that engaging multiple perceptual channels simultaneously can strengthen memory encoding and conceptual understanding [3, 5]. Haptic feedback has demonstrated efficacy in supporting procedural skill acquisition by conveying information through touch [2]. Since spelling fundamentally relies on procedural knowledge—the memorization of phoneme–grapheme correspondences and letter sequences [13]—there exists a compelling rationale for integrating tactile modalities into literacy instruction. Embodied cognition theory and dual coding theory further support this approach: bodily interaction with letters may establish action-based memory traces [6], while linking verbal stimuli with non-verbal imagery creates multiple retrieval pathways [8, 4].

Advances in generative AI have created opportunities for personalized content creation and real-time learner assessment, and research consistently demonstrates that digital learning tools achieve maximal efficacy when educators still remain central to the instructional process [11, 1]. This motivates systems that combine automated adaptation with teacher-facing analytics.

This paper introduces LexiLearn, an AI-enhanced immersive virtual environment designed for foundational literacy acquisition. The system synthesizes three technological streams: (1) an adaptive AI engine that continuously models learner proficiency and adjusts task parameters; (2) a multisensory virtual reality environment integrating visual, auditory, and haptic feedback; and (3) a teacher-facing AI-enhanced dashboard that translates learner analytics into actionable pedagogical insights.

2 Methods

This section describes how artificial intelligence was integrated throughout the LexiLearn system to enable adaptive activity generation, real-time learner modelling, and pedagogical support.

2.1 System Architecture Overview

LexiLearn comprises three interconnected subsystems: a VR learning environment built in Unity, an AI content and adaptation pipeline, and a web-based teacher dashboard that leverages natural language to suggest learner improvements. Fig. 1 illustrates the data flow among these components.

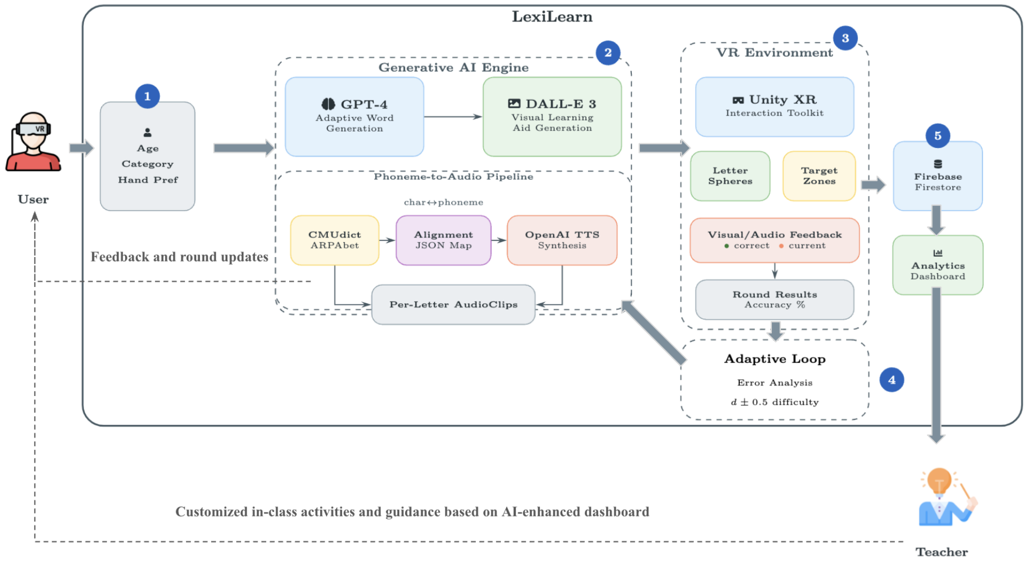

Fig. 1. System architecture showing data flow between the VR application, AI frameworks, Firebase backend, and Teacher Dashboard.

2.2 AI-Driven Adaptive Word Generation

The core adaptation mechanism employs GPT-4 to generate personalized word sequences in real time. Each API call transmits contextual information including the learner's age, current difficulty level (scaled 1–10), accuracy from the previous round, cumulative error patterns, and the selected word category. The model returns age-appropriate words calibrated to the learner's demonstrated proficiency.

The backend pipeline incorporates two priority modes. During the initial five words, an *age-based priority* emphasizes vocabulary and phonetic patterns typical for the learner's developmental stage. Subsequently, an *accuracy-based priority* shifts selection toward addressing identified weaknesses, with a 90% mastery threshold governing difficulty advancement. When learners consistently struggle with particular letter patterns (e.g., confusing "b" and "d", or omitting silent letters), the model generates targeted practice words featuring those patterns.

A progression safety mechanism prevents learners from becoming stuck at a single difficulty level. After three unsuccessful attempts to achieve 90% accuracy, the system advances the learner with a reduced difficulty increment (+0.25 rather than +0.5), ensuring continued engagement while maintaining appropriate challenge.

2.3 Generative Visual Content

DALL-E 3 produces child-friendly images representing each target word. The generation prompt explicitly requests imagery without embedded text, ensuring that visual associations support rather than substitute for orthographic processing. These images appear alongside the target word, establishing dual-coded verbal-visual linkages that support the development of orthographic-phonological-semantic mappings essential for fluent reading [16].

2.4 Phoneme-Grapheme Alignment Pipeline

A multi-stage pipeline handles the mapping between written letters and spoken sounds. The CMU Pronouncing Dictionary provides initial phoneme lookup for common English words. For words not in the dictionary or requiring grapheme-level alignment, GPT-4o-mini performs structured analysis, outputting JSON that maps each letter or letter cluster to its corresponding phoneme(s). This alignment enables per-character audio feedback synchronized with the learner's letter selections.

OpenAI's text-to-speech API generates audio for both whole-word pronunciation and individual phonemes. The per-phoneme audio clips are cached for efficiency and triggered in response to learner interactions, providing immediate auditory reinforcement of grapheme-phoneme correspondences.

2.5 Real-Time Data Collection and Learner Modelling

Every learner interaction is logged to Firebase Firestore in real time, creating comprehensive assessment records. Each round document captures anonymized user and session identifiers, category selected, difficulty level, round duration, character sequences entered versus target sequences, accuracy rate, and server timestamp.

The comparison of user entries against correct answers enables character-level error analysis, revealing specific letter confusions and systematic mistake patterns. These data inform the adaptive algorithm's subsequent word generation while simultaneously populating the teacher dashboard for longitudinal analysis.

2.6 AI-Enhanced Teacher Dashboard

The teacher dashboard translates learner activity within the VR environment into interpretable data supporting classroom instruction. Teachers can view session-by-session accuracy trends, words practiced, time-on-task metrics, and category-specific performance for each student.

A conversational AI assistant, powered by GPT-4 with access to learner performance data, enables teachers to query student progress using natural language. Teachers might ask questions such as "Which students are struggling with silent letters?" or "What activities would help Student 1 with her vowel confusion?" The AI analyzes error logs to identify specific literacy gaps—letter reversals, phonetic confusions, morphological errors—and generates evidence-based recommendations for classroom activities, including suggested materials, grouping strategies, and connections to curricular objectives.

3 Results

The integration of these AI components is designed to produce a cohesive immersive learning experience combining adaptive personalization with multisensory engagement.

3.1 The Learner Experience

The VR application situates learners within a nature-inspired virtual orchard designed to foster presence, curiosity, and sustained attention while minimizing visual distraction [9]. Learners navigate freely through the space, selecting from categorized word sets (animals, vehicles, emotions, food) that align with their interests and curricular objectives. Before entering gameplay, a customization menu allows learners to tailor aspects of their experience, fostering agency and intrinsic motivation [10]. (Link to demonstration video: `https://www.youtube.com/shorts/j29-6Xuvc5c`)

When a learner selects a category, a target word appears alongside an AI-generated image representing its meaning. Three-dimensional "fruit-letters" containing individual graphemes are presented in front of the learner. Each letter functions as an XR-interactable object that can be grasped, manipulated, and thrown into a collection basket. The learner's task is to select letters in the correct sequence to spell the target word, with more letters available than required to introduce productive challenge.

3.2 Multimodal Feedback Integration

Interaction with each letter triggers coordinated feedback across three sensory channels:

- **Visual feedback**: Letters change color upon selection (green for correct placement, red indicating the need to retry), with particle effects providing additional confirmation. The AI-generated image reinforces word meaning through dual coding.
- **Auditory feedback**: The corresponding phoneme is pronounced via text-to-speech synthesis, reinforcing grapheme-phoneme associations at the moment of interaction.
- **Haptic feedback**: Wearable gloves equipped with linear resonant actuators deliver localized vibrotactile stimulation to the hand and fingers. Each phoneme maps to a distinct vibration pattern, expanding grapheme-phoneme associations with a tactile dimension.

Upon correctly spelling the word, visual and auditory celebration cues precede the presentation of a new word, maintaining a continuous, process-oriented flow. The system deliberately avoids competitive scores or visible difficulty levels, promoting learning through exploration rather than performance pressure [12].

3.3 Adaptive Difficulty Progression

The AI engine maintains learners within their zone of proximal development [15] through continuous performance monitoring. As learners demonstrate mastery (achieving 90% accuracy), difficulty increases through longer words, more complex phonetic patterns, and reduced scaffolding. When learners struggle, the system responds by generating targeted practice addressing specific error patterns, providing individualized remediation without explicit intervention.

3.4 Teacher Dashboard Implementation

The web-based dashboard presents real-time proficiency tracking, per-student accuracy and completion metrics, and class-wide performance analytics. Common mistake identification highlights patterns across the classroom (e.g., frequent confusion between specific letter pairs), enabling teachers to address widespread challenges through whole-class instruction.

The conversational AI assistant positions insights as support for teacher professional judgment rather than prescriptive directives, respecting educator expertise while reducing the cognitive load of data interpretation. This design reflects the principle that AI systems should complement rather than replace human instructional decision-making.

4 Future Directions

The LexiLearn system has completed initial development with functional VR gameplay, AI content generation, Firebase data logging, and teacher dashboard interfaces. Ethical approval has been secured from the Institutional Review Board, and the project is entering its co-design and participant user-testing phase in partnership with local schools.

4.1 Planned Empirical Studies

Formative co-design sessions will iteratively refine the VR experience and dashboard based on teacher and student feedback. Classroom pilot studies with middle school classrooms will assess learning outcomes, engagement metrics, and teacher utility. Using standardized spelling and phonological awareness assessments, we will compare the multisensory VR condition against an audiovisual-only control condition to determine whether the addition of haptic feedback produces differential effects on spelling proficiency, phonological awareness, and lexical development.

Longitudinal analyses will examine whether the loop between automated adaptation and teacher-implemented instructional modifications produces cumulative benefits for literacy acquisition.

4.2 Research Questions

The planned studies will investigate several research questions:

1. Does the integration of haptic feedback with visual and auditory modalities enhance phoneme-grapheme association learning compared to audiovisual feedback alone?
2. What patterns of learner engagement and persistence emerge in the VR environment, and how do these relate to literacy outcomes?

4.3 Technical Extensions

Key design tradeoffs include caching phoneme audio to manage AI latency, hiding scores to reduce performance anxiety, and a three-attempt safety mechanism for stuck learners. Classroom deployment considerations include session duration limits, anonymized IDs for FERPA-compliant data privacy, and adjustable haptic intensity for sensory-sensitive learners.

A classroom code system in development will enable teachers to generate unique identifiers linking VR session data to class rosters, facilitating real-time progress monitoring.

Future development will explore multimodal analytics (gaze tracking, gesture recognition), expansion to additional educational domains, and deployment of distilled models on VR headsets for reduced latency and offline capability.

References

[1] Campen, K. V., Molenaar, I.: How Teachers Integrate Dashboards into Their Feedback Practices. *Frontline Learning Research* **8**(4), 37–51 (2020)

[2] Culbertson, H., Schorr, S. B., Okamura, A. M.: Haptics: The present and future of artificial touch sensation. *Annual Review of Control, Robotics, and Autonomous Systems* **1**(1), 385–409 (2018)

[3] Damsgaard, L., Nielsen, A.-M. V., Gejl, A. K., Malling, A. S. B., Jensen, S. K., Wienecke, J.: Effects of 8 weeks with embodied learning on 5-6-year-old Danish children's pre-reading skills and word reading skills: The PLAYMORE Project, DK. *Educational Psychology Review* **34**(3), 1709–1737 (2022)

[4] Ernst, M. O., Banks, M. S.: Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**(6870), 429–433 (2002)

[5] Fredens, K.: *Læring med kroppen forrest: En hjernebrugsbog*. Akademisk Forlag (2018)

[6] Glenberg, A. M., Kaschak, M. P.: Grounding language in action. *Psychonomic Bulletin & Review* **9**(3), 558–565 (2002)

[7] Oakley, G.: A scoping review of research on the use of digital technologies for teaching reading fluency. *Education Sciences* **14**(6), 633 (2024)

[8] Paivio, A.: *Mental Representations: A Dual Coding Approach*. Oxford University Press (1990)

[9] Queiroz, A. C. M., Nascimento, A. M., Tori, R., da Silva Leme, M. I.: Using HMD-based immersive virtual environments in primary/K-12 education. In: International Conference on Immersive Learning, pp. 160–173. Springer International Publishing, Cham (2018)

[10] Ryan, R., Deci, E.: Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist* **55**, 68–78 (2000)

[11] Selwyn, N.: *Education and Technology: Key Issues and Debates*. Bloomsbury Publishing (2021)

[12] Skovbjerg, H. M., Bekker, T.: *The Value of Play*. Design School Kolding (2018)

[13] Steffler, D. J.: Implicit cognition and spelling development. *Developmental Review* **21**(2), 168–204 (2001)

[14] UNESCO: *UNESCO Strategy for Youth and Adult Literacy (2020–2025)* (40 C/25). United Nations Educational, Scientific and Cultural Organization (2019)

[15] Vygotsky, L. S.: *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press (1978)

[16] Warmington, M., Hulme, C.: Phoneme awareness, visual-verbal paired-associate learning, and rapid automatized naming as predictors of individual differences in reading ability. *Scientific Studies of Reading* **16**(1), 45–62 (2012)