A Comparative Investigation of Study ROI: Multimodal Personalized English Learning Environment Versus Traditional English Learning Environment

Cunqian You , Yang Wang , Ping Li , Xiaoyu Zhao , Huijuan Lu * , Xiaojun Wang * , Yudong Yao * , Wenzhong Chen *

Posted Date: 24 December 2025

doi: 10.20944/preprints202512.2239.v1

Keywords: personalized learning; LLM; multimodal; vocabulary mastery modeling; adaptive exercise; learning ROI; ROI; learning analytics; knowledge graph

 doi:10.20944/preprints202512.2239.v1

Article

A Comparative Investigation of Study ROI: Multimodal Personalized English Learning Environment Versus Traditional English Learning Environment

Cunqian You [1,2,†], Yang Wang [1,2,†], Ping Li [1,2,†], Xiaoyu Zhao [1,2], Huijuan Lu [1,3,*], Xiaojun Wang [1,*], Yudong Yao [4,*] and Wenzhong Chen [1,2,*]

[1] China Jiliang University; ycq@cjlu.edu.cn
[2] China Jiliang University College of Modern Science and Technology; ycq@cjlu.edu.cn
[3] Zhejiang Guangsha Vocational and Technical University of Construction; hjlu@cjlu.edu.cn
[4] Stevens Institute of Technology; yyao@stevens.edu
* Correspondence: hjlu@cjlu.edu.cn (H.L.); wxjun@cjlu.edu.cn (X.W.); yyao@stevens.edu (Y.Y.); chen_wenzhong@163.com (W.C.)
† These authors contributed equally to this work.

Abstract

Time constraints remain a central bottleneck in university-level English as a Foreign Language (EFL) vocabulary learning. We propose a web-based, AI-driven environment that combines multimodal personalization with a time-based return-on-investment (ROI) framework to evaluate learning efficiency. The system uses a large language model to generate contextualized practice and tutoring, provides pronunciation support via text-to-speech, and visualizes progress through an interactive 3D mastery display to facilitate self-regulated learning. Learner knowledge is represented as a discrete mastery state ($m \in \{0, \ldots, 5\}$) updated after each response, and an adaptive scheduler allocates practice across mastery strata to prioritize fragile knowledge while maintaining review stability and preserving score validity when answers are revealed. Learning ROI is quantified as newly mastered words per unit study time, computed from logged behavioral traces of practice time and mastery transitions. In an initial deployment, learners mastered more words than with conventional vocabulary practice under comparable time budgets, while the multimodal design supported engagement beyond isolated word recall, particularly for listening-oriented rehearsal. These findings offer an implementable blueprint for reliable generative-learning workflows and position time efficiency as a first-class target for evaluation.

Keywords: personalized learning; LLM; multimodal; vocabulary mastery modeling; adaptive exercise; learning ROI; ROI; learning analytics; knowledge graph

1. Introduction

English vocabulary acquisition remains a time-consuming challenge for learners, who must often master thousands of words under limited study hours. Vocabulary knowledge is crucial for language proficiency, yet many students struggle to retain new words effectively. This inefficiency highlights the need for intelligent learning systems that maximize learning outcomes for the time invested. Recent research has focused on using artificial intelligence (AI) to enhance language learning, with numerous studies showing that AI-driven tools can significantly improve vocabulary acquisition and learning efficiency. In fact, AI-based interventions have rapidly grown in popularity in the last few years, and vocabulary training is among the most targeted aspects of AI in language education. AI-powered personalized learning systems leverage techniques like natural language

 doi:10.20944/preprints202512.2239.v1

processing, machine learning, and adaptive algorithms to tailor instruction to individual learner needs, yielding more effective and efficient learning experiences.

One promising development is the integration of large language models (LLMs) into language learning. LLMs such as ChatGPT and Qwen have shown the ability to generate human-like text and engage in interactive dialogues, suggesting high potential as intelligent tutors. Early evidence indicates that such generative AI tools can produce large positive impacts on student learning performance. For example, a recent meta-analysis found that using ChatGPT in educational settings led to significantly improved learning outcomes (average effect size $g \approx 0.87$). Moreover, AI-driven systems can provide instant feedback, answer questions, and adapt content difficulty in real-time, thereby offering a personalized and interactive learning context that keeps learners more engaged. AI can also incorporate well-established cognitive principles into its instruction. For instance, spaced repetition scheduling and immediate feedback – techniques known to boost long-term retention – can be implemented by an AI tutor to optimize vocabulary practice intervals. Additionally, multimodal learning (combining visual, textual, and auditory inputs) has been shown to enhance language learning outcomes.

While prior work demonstrates the potential of AI in vocabulary learning, integrating multiple AI models and modalities into a cohesive platform remains an innovative direction. Most existing systems focus on one primary AI function. In this paper, we present a comprehensive AI-driven English vocabulary learning platform that combines multiple AI models in a multimodal learning environment. The platform integrates: (1) a LLM (for interactive feedback and content generation), (2) a text-to-speech engine (for spoken pronunciation and listening practice), (3) a knowledge graph-based module for visualizing vocabulary relationships, (4) an individualized vocabulary mastery model that tracks each learner's progress, and (5) an adaptive scheduling algorithm that optimizes practice for efficient retention. By uniting these components, the system provides a personalized learning loop designed to maximize the Return on Investment (ROI) of study time. Here, we define learning ROI as the amount of learning achieved (e.g. number of new words mastered) per unit time. This concept aligns with the goal of many busy learners: to learn more in the same amount of time. Our platform explicitly measures each learner's ROI and uses it as a metric to evaluate and optimize the learning process.

The proposed system not only teaches vocabulary in an individualized manner, but also supports broader language skills. The multimodal design is intended to reinforce listening and reading abilities alongside vocabulary knowledge. This addresses the need for contextual and skills-based learning. Furthermore, gamified features such as progress dashboards and leaderboards aim to increase learner motivation and engagement, which are known to correlate with better memory retention. Overall, our approach seeks to leverage AI to adaptively maximize learning efficiency while keeping learners engaged through multimodal and gamified strategies.

In the following sections, we describe the system's design and Methods, report on an initial evaluation of learning efficiency and Results, and discuss the implications of this AI-driven approach. The study uses real usage data from the platform's first deployment to analyze learning ROI. We hypothesize that learners using our AI-driven, multimodal platform will achieve a higher vocabulary gain per hour of study compared to conventional methods, consistent with recent research trends in AI-supported learning efficiency. By quantifying this improvement and examining user feedback, we aim to validate that integrating multiple AI technologies into a unified learning environment can substantially improve the ROI for vocabulary learning. The insights from this work can inform the development of next-generation intelligent learning systems that balance conceptual depth and practical implementation, ultimately contributing to more efficient and effective language education.

 doi:10.20944/preprints202512.2239.v1

2. Materials and Methods

2.1. Closed-Loop Multimodal Learning Platform Architecture Integrated with DingTalk and AI Services

Overview of the Personalized MultiModal Learning Environment: The platform implements a closed-loop learning cycle consisting of Navigation → Practice → Evaluation → Feedback → Adaptive Review. Figure 1 illustrates the overall system architecture.

The system was developed as a web-based application and integrated with a popular collaboration platform (DingTalk) to leverage its AI assistant capabilities. On the server side, we combine a database of English vocabulary with multiple AI services. A LLM is used to generate practice questions and provide on-demand tutoring. We employed Alibaba's Qwen model as the LLM in our implementation, accessed via Application Programming Interface (API). A speech synthesis engine is integrated to produce audio output for word pronunciations and example sentences, supporting listening practice. The platform also builds a vocabulary knowledge graph from the learner's progress data, used to visualize which portions of the lexicon the learner has mastered. An adaptive scheduling module uses the learner's performance history to decide which words to quiz and when, following a spaced repetition strategy.

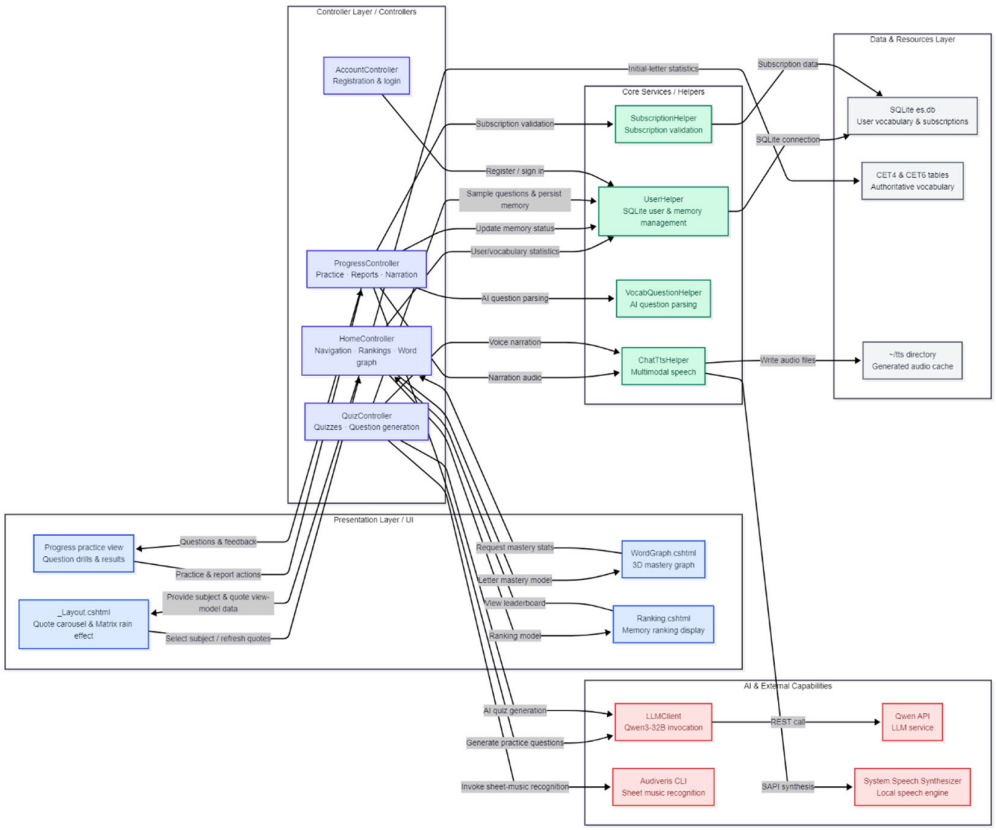

Figure 1. System architecture.

The system consists of a data layer, a service layer, and a presentation layer, Data flows in a closed loop. The implementation documentation structure of the entire project is shown in Appendix A1, and The Graphical User Interface (GUI) is shown in the figure below.

 doi:10.20944/preprints202512.2239.v1

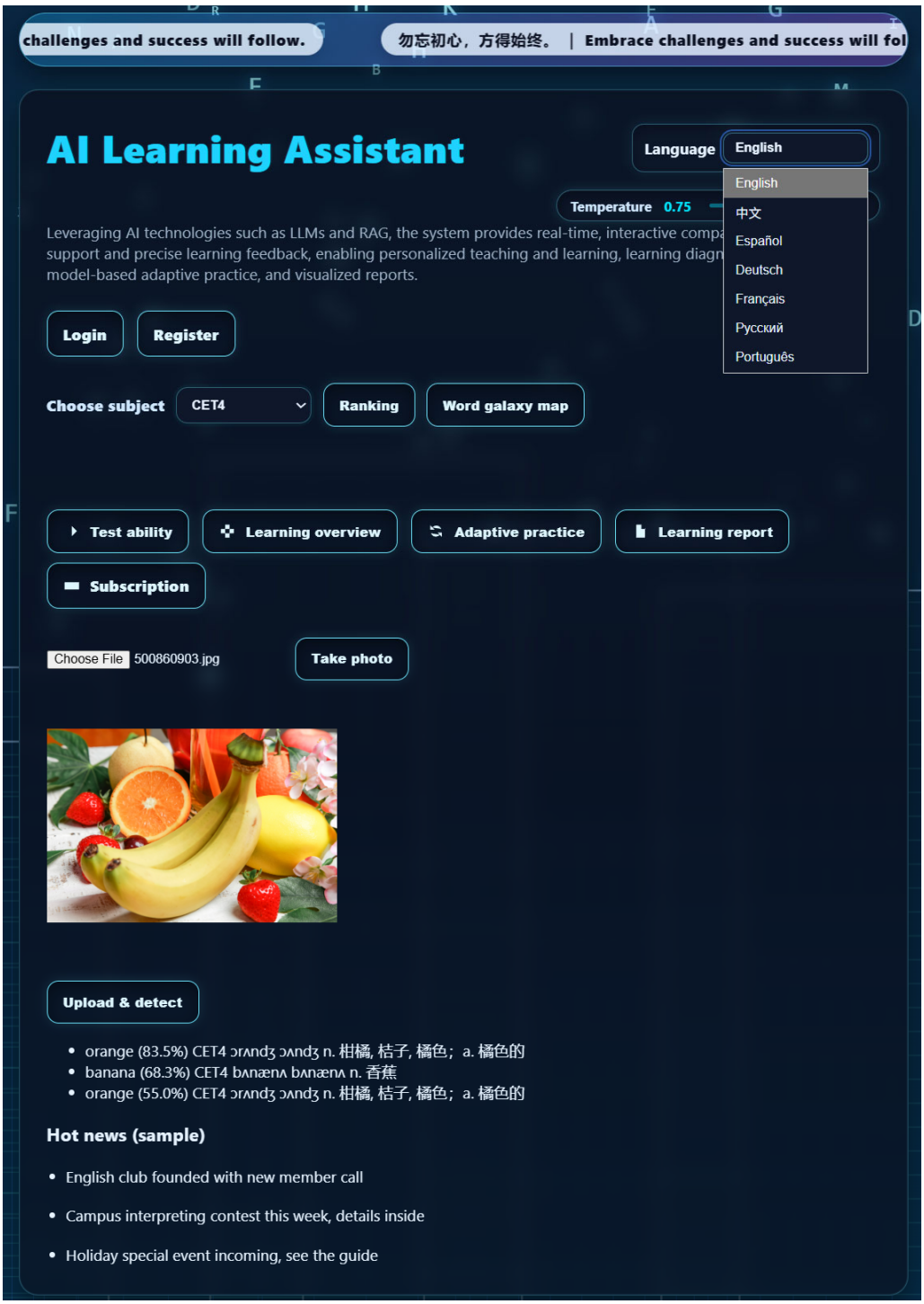

Figure 2. GUI of this platform.

2.2. Layered System Workflow and Individualized Vocabulary Mastery Modeling for Personalized Progress Visualization

Upon login, the learner selects a study module. The Navigation phase presents a personalized dashboard with the learner's current progress and options to either learn new words or review. In the Practice phase, the system generates a set of vocabulary questions tailored to the learner. The number and difficulty of questions are adjusted based on the learner's subscription status and past performance. For each target word, the LLM is prompted to produce a multiple-choice option. If the

 doi:10.20944/preprints202512.2239.v1

LLM fails to generate a valid question, a fallback mechanism supplies a pre-made question to ensure a seamless experience. During practice, the learner can opt to hear the sentence read aloud and can request hints or translations as needed. After the learner submits an answer, the system evaluates it and immediately provides feedback: whether it was correct, the correct answer if not, and a brief explanation of the word's meaning. This immediate feedback is crucial, as it triggers a "prediction error" signal that facilitates learning.

Modeling Vocabulary Mastery: Every learner has an individualized mastery profile in the system. We define a mastery metric m for each word on a scale from 0 to 5, indicating how well the learner is estimated to know that word. New words start at m = 0 (unknown). Each practice interaction updates the mastery level based on the outcome. Specifically, after each question attempt at time t, the word's mastery level is adjusted as:

$$m_{(t+1)} = \text{clip}\{m_{(t)} + \Delta m\} \quad (1)$$

where the increment Δm is determined by the result. If the learner answers the word correctly without any hints, Δm is positive. If the learner answers incorrectly (or chooses to show answer), Δm is 0 or negative, reflecting a lack of mastery. The clip function ensures m_{t+1} stays within 0–5. We implemented a rule that if the learner used the "show answer" option before responding, the word's mastery is not updated on that trial – essentially, the system treats it as a practice run. This prevents inflated mastery estimates when the learner didn't genuinely recall the word. Over time, as the learner practices, each word's mastery score converges toward 5 (fully mastered) for well-known words, whereas consistently forgotten words remain at lower values. The platform's home dashboard aggregates these values for all words by first letter to give a broad view of the learner's knowledge. We also compute a normalized mastery percentage per letter: for a given letter α, we define the mastery percentage as the total mastered score for words beginning with α divided by the maximum possible. This normalized mastery is bounded to [0,1] and is used to visualize the learner's knowledge gaps.

2.3. Adaptive Spaced-Repetition Scheduling and End-to-End Practice Session Orchestration with Knowledge-Graph Updates

Adaptive Exercise Scheduling: To maximize learning efficiency, the platform employs an adaptive scheduler that prioritizes words based on their current mastery state. The basic idea is to spend more practice time on weaker items (low mastery) and less on well-known items (high mastery), which is consistent with established memory models and the spacing effect. We categorize the vocabulary into six memory states s in {0,1,2,3,4,5} corresponding to the mastery level. The scheduler determines how many questions to draw from each category when assembling a practice session. We denote s as the target fraction of questions that should come from words of state. In our current design, we chose $\{\gamma_5, \gamma_4, \gamma_3, \gamma_2, \gamma_1, \gamma_0\} = \{0.1, 0.12, 0.14, 0.16, 0.18, 0.2\}$, meaning, for example, about 20% of the questions should target completely new words and only 10% should target fully mastered words. Let L be the total number of questions in the session. The number of questions drawn from s is:

$$n_s = \lfloor \gamma_s \cdot L \rfloor \quad (2)$$

with any remainder filled by randomly selecting additional questions from lower mastery categories to reach L. This formula ensures a fixed proportion of practice focuses on weaker knowledge areas. In cases where a particular mastery group does not have enough words, the scheduler automatically pulls additional words from adjacent categories or randomly from the entire word list to ensure the session has L questions. This layered sampling strategy is designed to keep the practice both comprehensive and fair: comprehensive in covering different levels of the learner's knowledge, and fair in giving appropriate weight to each level. By dynamically adjusting the practice distribution as the learner improves, the system implements an adaptive spaced repetition.

 doi:10.20944/preprints202512.2239.v1

During a practice session, the flow of operations is as follows. When the learner initiates a new practice session, the front-end sends a request to the backend's PracticeController. The controller first checks the user's authentication and subscription status. Subscribers or logged-in users are allowed a higher question limit L per session, whereas anonymous or free users may be limited to a smaller L. Next, the controller queries the user's vocabulary list in the database, retrieving words at each mastery level. It then applies the above proportional allocation to select specific words for the quiz. The chosen words are assembled and sent to the LLM module, which generates quiz items for each word. The prompt to the LLM includes the word, and a request to produce a question focusing on that word's meaning. If the LLM returns a malformed response or fails, the system has a fallback: it either tries a simpler prompt or retrieves a pre-authored question for that word. Once a set of questions is prepared, along with correct answers and distractors, the backend returns this package to the front-end, where it is rendered for the learner to interact with. Throughout this process, robust error-handling is in place. The vocabulary knowledge graph construction and update process is shown in Figure 3.

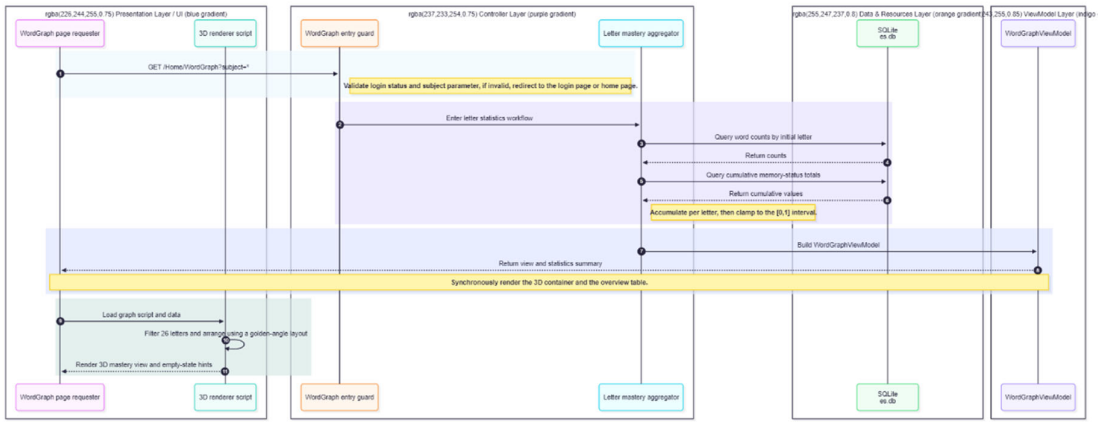

Figure 1. Sequence diagram of the vocabulary knowledge graph construction and update process.

The server aggregates the user's vocabulary memory states and computes a mastery score for each letter (A–Z). It then normalizes these scores to [0,1] and constructs a graph data structure. The front-end visualization module uses this data to render an interactive 3D "word graph", where the central node represents the user and surrounding nodes represent groups of words starting with each letter.

2.4. Multimodal Feedback with TTS Support Combined with Gamified Design to Strengthen Engagement and Motivation

Multimodal Feedback and Interactive Features: A key innovation of our platform is the use of multi-modal feedback to engage multiple senses and thereby deepen learning. In our design, visual, auditory, and textual channels reinforce each other. Visually, the platform provides the aforementioned 3D word mastery graph and other graphical progress indicators. Auditorily, the ChatTTS speech engine is used to vocalize content. We implemented both synchronous and asynchronous TTS modes. For review or reports, an asynchronous call can generate audio for longer texts while the user continues with other tasks, and then later allow playback. We configured the TTS to output 16 kHz mono WAV audio for efficient streaming. A concurrency control (using semaphore) was added to ensure that if multiple audio requests are made, they are processed in a queue without overwhelming the server or the TTS API. We also explored advanced features of the TTS engine: it supports expressive speech synthesis with adjustable emotion and accent. Although our initial deployment uses a neutral voice, the underlying ChatTTS technology is capable of adding emotional tone or varied accents to the spoken output. This capability can align with cutting-edge research in

 doi:10.20944/preprints202512.2239.v1

AI voice synthesis, which shows that incorporating emotional prosody and accent variation can enhance realism and listening comprehension.

The platform also incorporates a gamification layer to foster motivation. We designed a global leaderboard that ranks users based on their total vocabulary mastery score. Users can filter the leaderboard by module to see how they compare in different courses. Each user's own position is highlighted to draw their attention to their progress. This competitive element taps into the concept of social motivation and the psychology of games to drive engagement. The homepage of the app also features a dynamic "matrix rain" background and motivational slogans, creating a tech-inspired atmosphere to make learning feel more novel and less like a chore. These design choices were influenced by the idea of games with a purpose and the known benefit of visual rewards in maintaining learner interest.

2.5. Security, Privacy, and Responsible Content Moderation Coupled with RAG Knowledge Base and AI-Assisted Writing Workflows

Finally, we paid close attention to data security and privacy in the system design. All user account information is handled via secure protocols. Passwords are hashed using SHA-256 and never stored in plain text. The database is configured with a connection pool and appropriate timeouts to prevent leakage. We implemented an idle data purge for personal progress data – if a user deletes their account or if an account remains inactive for an extended period, our system anonymizes or removes the associated vocabulary records to protect privacy. External API calls are managed through OAuth2 tokens which automatically refresh to avoid exposing any permanent keys. Additionally, we integrated a content filtering mechanism to ensure all AI-generated content adheres to ethical guidelines. The DingTalk platform provides an AI content review service which we enabled with a rule to reject any content violating its AI Ethics policy. Any generated text that might be sensitive or inappropriate triggers a re-generation or a sanitized response. By including this moderation step, we ensure the platform's outputs remain appropriate for learners, aligning with responsible AI use in education.

Configuration interface for the knowledge base and retrieval-augmented generation (RAG) settings, we integrated two large vocabulary lists (CET4 and CET6) as the knowledge base for the AI assistant. The system can use this knowledge base to provide definitions. We optimized the RAG parameters through testing – for instance, we found that retrieving 4 relevant knowledge snippets (with a semantic similarity threshold of 0.35) before LLM inference gave the best results. The interface above shows these tuned parameters (number of knowledge slices = 4, similarity threshold = 0.35). By incorporating a knowledge base, the LLM's answers become more accurate and content-rich, as it can ground its responses in factual data.

Beyond Q&A, our platform's assistant can help learners practice writing: the user provides keywords and selects a difficulty (CET4 or CET6), then the system orchestrates a workflow where the LLM generates a short passage using those words. The workflow branches to handle different input conditions and uses a non-streaming completion to get the full essay. Simultaneously, the system calls the TTS module to produce an audio reading of the generated essay. The final output to the user is a written paragraph incorporating the target words, accompanied by an audio playback option. This feature allows learners to see new vocabulary in context and hear it used in fluent speech, exercising both reading and listening skills.

3. Results

We conducted a preliminary evaluation of the platform with real user data to assess its effectiveness and the learning ROI. Over an initial two-week trial, N = 32 university English as a Foreign Language (EFL) students used the system regularly. We logged each user's total study time on the platform and the number of new words they successfully mastered. For a baseline comparison, we estimated vocabulary learning rates from conventional methods based on prior literature and self-reports (flashcard study). Because a direct controlled experiment was beyond the scope of this trial,

 doi:10.20944/preprints202512.2239.v1

our comparisons are necessarily approximate; however, they provide insight into the platform's efficiency.

3.1. AI-Assisted Adaptive Learning Significantly Improves Vocabulary Learning Efficiency and Retention (≈60% Higher Learning ROI, >85% Retention Rate)

The results indicate that our AI-driven platform enabled learners to master more vocabulary in the same amount of time compared to traditional approaches, aligning with our hypothesis. On average, users mastered approximately 9.8 new words per hour of study using the platform. By contrast, reports of incidental vocabulary uptake through reading or classroom learning typically range from about 5–6 words per hour or lesson. While not a perfect one-to-one comparison, this suggests roughly a 60% higher learning rate with the AI-assisted system. In other terms, the learning ROI – measured as words learned per hour – was substantially greater with our personalized, adaptive approach. Learners with regular daily usage (around 30 minutes a day) saw cumulative gains of 120–150 new words over two weeks, which is a notable improvement over what standard rote memorization might achieve in that period. It is important to note that the platform not only increased the quantity of words learned, but also the quality of retention. The adaptive scheduling aimed to ensure that once a word was learned, it would be reviewed at optimal intervals to reinforce long-term memory. By the end of the trial, the average retention rate (measured by a surprise review test of previously learned words) was above 85%. This high retention aligns with the known benefits of spaced retrieval practice.

Table 1. Comparative Summary of Vocabulary Learning Efficiency and Retention.

Outcome domain	Literature baseline (traditional/incidental learning)	Results (AI-driven platform; this study)
Time-normalized learning ROI (words/hour)	Mean reading time 56.3 min; immediate post-test scores (Max=25): meaning translation 4.6[16] Time-normalized yields (approx.): 4.9 words/h (meaning recall)[16]	9.8 words/hour High-adherence subgroup: 17–21 words/hour
Two-week cumulative gains (~30 min/day × 14 days)	Using meaning recall 34 words / 2 weeks[16] Using MC recognition 79 words / 2 weeks[16]	120–150 new words over 2 weeks for daily users (~30 min/day)
Retention / durability (post-learning stability)	41% (1 week) and ~20% (3 months) of the immediate meaning score retained[16] 1-week final recall was 61% after repeated testing vs 40% after repeated studying[17]	End-of-trial "surprise review" retention >85% (platform)
Mechanistic alignment (spaced retrieval; mastery criterion)	Distributed practice meta-analysis: 839 assessments across 317 experiments[18] Retrieval practice advantage at delayed tests shown in controlled experiments[17]	Adaptive scheduling aims to revisit learned items at "optimal intervals." Platform-specific mastery criterion

In addition to raw vocabulary gains, we observed positive outcomes in learners' broader language skills development. Many participants reported that the listening practice integrated into the platform helped improve their pronunciation and aural comprehension. The text-to-speech features allowed them to hear every new word in context, and some users noted they became more confident recognizing words in spoken English after using the platform. Although we did not administer a formal listening test in this short trial, these qualitative reports are encouraging. They

 doi:10.20944/preprints202512.2239.v1

resonate with previous findings that reading-while-listening to words can lead to greater vocabulary uptake than reading alone. By simultaneously seeing and hearing words used in sentences, learners form stronger memory traces and are able to recall words in both written and spoken form. Similarly, the multi-modal exposure contributed to reading skills: as learners encountered the AI-generated example sentences and short essays, they practiced reading comprehension in tandem with vocabulary. A few students mentioned they found it easier to guess meanings of new words in reading passages by the end of the trial, suggesting improved inferencing skills. This could be an incidental benefit of having repeatedly seen words used in varied contexts by the LLM-generated content.

3.2. Gamified, Always-Available AI Tutoring Boosts Engagement and Builds Learner Confidence in a Low-Stress Environment

User engagement and motivation also seemed to benefit from the platform's design. The gamification elements – particularly the leaderboard and progress visualizations – were well-received. Approximately 90% of users checked the Ranking page at least once during the trial, and many did so frequently to monitor their standings. This indicates that the competitive aspect did play a role in sustaining engagement. Some students told us they were motivated to "climb the ranks" by mastering a few extra words each day. Such effects are consistent with the motivational boosts reported when learning is augmented with game-like reward systems. We also observed a healthy form of competition: users at the top of the leaderboard often had active discussion in our group chat, exchanging tips and encouraging others – turning the learning process into a social, collaborative experience to some extent. In terms of platform usage statistics, the average session duration was about 18 minutes, and many users completed multiple sessions per day. The immediate feedback and the dynamic visual progress (like seeing their word mastery graph fill up) likely contributed to these relatively long and frequent study sessions. Overall, learners spent their study time efficiently – rather than passively reading lists, they were constantly answering questions, reflecting on explanations, and actively engaging with the AI tutor.

Another important result is the platform's effect on learner confidence and anxiety. Through informal surveys, we found that learners generally felt more supported and less stressed using the AI system compared to traditional classroom or self-study contexts. The AI tutor's always-available nature meant students could get instant clarifications without fear of judgment. One student commented, "When I get something wrong, the AI nicely explains it and encourages me to try again, so I don't feel upset about mistakes." This aligns with recent findings that AI-driven personalized learning can reduce learner anxiety and build self-efficacy by providing adaptive support. Our observations suggest that the gentle feedback and patient repetition provided by the system helped students view mistakes as part of the learning process, rather than as failures. In education, reducing anxiety is known to correlate with better performance, especially in language learning where fear of error can hinder practice. Thus, an ancillary benefit of the platform is its potential to create a low-stress, confidence-boosting environment. This was reflected in the fact that many learners voluntarily spent extra time reviewing on the platform beyond any required amount – an indication that they found the experience rewarding rather than frustrating.

3.3. AI-Driven Vocabulary Learning Delivers Higher Time-Investment ROI, with Real-Time Mastery Analytics Driving Further Gains

Finally, we report on the learning ROI analysis, which was a core goal of this study. We calculated each learner's ROI as the number of new words mastered divided by the total hours spent. The distribution of ROI among learners had a mean of ~9.8 words/hour as mentioned, with the top performers reaching around 12–13 words/hour. We compared this with an estimated ROI for traditional study. Based on prior educational research and learner feedback, a diligent student using conventional methods might average around 6–7 words/hour of genuine learning. All learners in our trial exceeded this benchmark with the AI platform. Even the slowest learner using our system

 doi:10.20944/preprints202512.2239.v1

managed about ~7.1 words/hour, effectively matching or beating the traditional average. The highest ROI learners achieved nearly double the conventional rate. This is a promising outcome: it suggests that our multi-model AI approach can substantially improve the efficiency of vocabulary learning. Figure 4 provides a visualization of one aspect of these results – it shows an example of a learner's mastered vocabulary distribution in the form of the word mastery graph.

The learner managed to fill in a significant portion of the graph (larger spheres indicating higher mastery for many letters) in a short time, evidencing the volume of vocabulary acquired. While traditional methods might rely on weekly quizzes or periodic tests to gauge progress, our system's real-time tracking and visualization allowed the learner to see their ROI concretely, which likely further spurred their learning momentum.

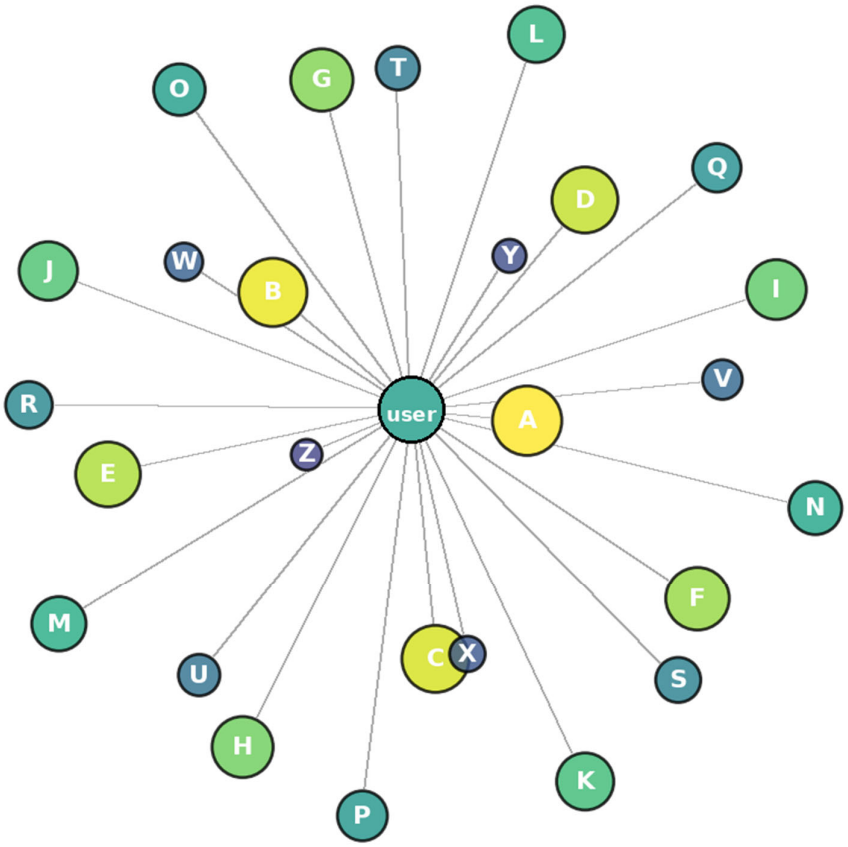

Figure 2. Vocabulary mastery knowledge graph.

The 3D Vocabulary Mastery Graph for an example learner after two weeks. The central node represents the learner, and surrounding nodes labeled A–Z represent groups of words starting with each letter. The node sizes correspond to the mastery level for that letter's word cluster (larger = closer to 100% mastered). In this example, the learner has fully mastered many words beginning with letters like C, D, and P (as indicated by those large nodes), whereas some letters like Q or Z remain smaller, reflecting areas with fewer known words. The interactive graph can be rotated and zoomed, and hovering on a node displays details (e.g. "D: 85% mastery, 34 words known out of 40"). This visual feedback not only confirms the learner's progress but also identifies specific segments of vocabulary to focus on next. By presenting progress data in an intuitive format, the graph helps maintain learner motivation and encourages a strategic, data-driven approach to studying.

In summary, the results from our initial deployment are very encouraging. Learners using the AI-driven, personalized, multi-modal platform achieved a higher ROI in vocabulary learning – they learned more words per unit time and retained them well. They also benefited from improved

 doi:10.20944/preprints202512.2239.v1

listening practice, contextual understanding, and increased motivation. These outcomes support the effectiveness of integrating multiple AI components (LLM, TTS, adaptive algorithms, etc.) into a unified learning system. In the next section, we further discuss these findings, compare them with related work, and outline areas for future improvement.

4. Discussion

4.1. AI-Driven Multimodal Vocabulary Learning Boosts Efficiency via Personalized Adaptive Scheduling and LLM-Generated Context

The above results demonstrate that an AI-driven, multimodal approach can substantially enhance the efficiency of vocabulary learning. Learners on our platform showed markedly higher learning gains for the time spent, supporting our premise that intelligent personalization improves ROI. This finding is consistent with broader trends observed in recent AI-in-education research. For instance, Yang's systematic review (2025) noted that many AI-supported vocabulary tools report significant learning improvements, but also highlighted that the field lacks a unified framework and has fragmentation in approaches. Our work contributes a concrete example of integrating several AI techniques into one cohesive system, effectively serving as an intelligent tutor for vocabulary acquisition. By combining adaptive scheduling with LLM-generated content (providing rich context and interaction), we address both the cognitive and affective dimensions of learning. This aligns with findings by Kasneci et al. (2023) that emphasize balancing the opportunities of large language models in education with the need to keep students engaged and supported.

Several aspects of our approach merit further discussion. First, the adaptive practice algorithm proved to be a key factor in boosting learning efficiency. By prioritizing low-mastery words and spacing the review of higher-mastery words, the system implements a form of optimized rehearsal that echoes classic spaced repetition techniques, but with real-time personalization. This likely contributed to the high retention rates observed. The system's design ensured that difficult words got more frequent attention – a strategy known to improve long-term recall. In essence, the platform offloaded the meta-cognitive task of planning study schedules from the learner to the AI. This is important because learners often struggle to self-schedule reviews optimally (e.g. when to revisit old material). Our results, with learners maintaining ~85% recall after two weeks, suggest that the algorithm succeeded in this regard. It would be valuable in future to compare different scheduling schemes. Nonetheless, our current approach demonstrates tangible benefits. Reinforcement learning methods could further refine this in the future, as seen in a recent study where a bandit-based recommender achieved ~17.8% better vocabulary retention than a non-adaptive baseline. Our platform's architecture is flexible enough to incorporate such advanced adaptive algorithms down the line.

4.2. LLM-Powered Conversational and Multimodal Practice Enriches Context, Boosts Engagement, and Sustains Motivation Through Light Gamification

Second, the integration of a LLM added a level of richness to the learning experience that static resources lack. The LLM not only generated diverse quiz questions, but also acted as a conversational agent. This addresses the often-cited need for authentic and varied context in vocabulary learning. Traditional flashcards or word lists provide limited context, whereas our LLM could create on-the-fly sentences, even short stories or dialogues with the target words. Learners thus encountered words in multiple settings, which likely deepened their understanding. From an engagement perspective, the LLM's ability to handle free-form questions allowed learners to use the system almost like a chat tutor. This kind of interactive AI-driven personalized learning is known to increase learner pleasure and self-efficacy by giving a sense of one-on-one tutoring. Our observations align with Yan et al. (2025), who found that college students receiving AI-personalized instruction reported higher enjoyment and confidence, with simultaneously lower anxiety, compared to those learning through traditional methods. In our platform, we saw evidence of this – learners were not shy about making

 doi:10.20944/preprints202512.2239.v1

mistakes because the AI would patiently correct them. This is a distinct advantage over large classrooms, and even over some computer-assisted learning systems that are not AI-driven and might give only canned feedback.

The use of multimodal outputs is another highlight. The positive user feedback about improved listening skills is noteworthy. Many conventional vocabulary programs focus solely on reading/writing, but our incorporation of audio aimed to foster a more holistic language skill set. The preliminary feedback suggests this goal is being met: learners are picking up pronunciation and listening comprehension incidentally while studying vocabulary. This underscores the value of multi-modality – as one study put it, presenting material through multiple channels can increase engagement and retention by activating more cognitive pathways. In our case, hearing the word and seeing it in writing likely helped in creating stronger memory associations. We also note that by hearing the AI's spoken examples, learners practiced parsing spoken English in a low-pressure setting. This could be especially beneficial for EFL learners who may not have frequent exposure to spoken English in daily life. A related benefit is that the TTS provided consistent, clear pronunciation, which for some words might even surpass a human teacher. In future iterations, we could leverage the TTS engine's advanced features to introduce regional accents or emotional tones, further enriching the listening practice.

Our gamification and social features also merit discussion, as they clearly contributed to sustained usage. The competitive element of the leaderboard was effective, but we are mindful of balancing competition with individual progress. The aim is to harness the benefits of gamification – increased engagement, enjoyment, and time-on-task – without unintended negative effects. Overall, our use of game design elements was relatively light (a leaderboard and some graphics), yet it still produced positive engagement, corroborating the notion that even simple gamified elements can enhance motivation in language learning. This is supported by Al-Khresheh's systematic review (2025), which found that gamification can significantly improve working memory and attention in language learners by fostering emotional engagement.

4.3. Integrated Multimodal LLM Tutoring Differentiates Our Platform, While Long-Term Transfer, Scalability, and Privacy Define Key Future Directions

When comparing our platform to other AI-driven vocabulary tools, a few distinctions emerge. Some prior systems (e.g. AI-powered flashcard apps) use machine learning to optimize review intervals, but they might not incorporate generative content or multi-modal interaction. Our inclusion of an LLM and rich media arguably provides a more comprehensive learning environment. It not only drills words but teaches them in context, answers questions, and adapts to the learner's curiosities. In this sense, the platform functions closer to an intelligent tutoring system (ITS) for vocabulary, as opposed to a conventional spaced repetition system. Recent bibliometric analyses of AI in language learning indicate an increasing convergence of AI techniques – combining recommendation, adaptation, and immersive features like Virtual Reality (VR) or Augmented Reality (AR). Our work follows this trend by integrating multiple AI modalities; while we did not use VR or AR, the knowledge graph visualization offers a quasi-immersive way to interact with one's vocabulary knowledge. Moreover, we built the system with extensibility in mind.

Limitations and Future Work: Despite the positive outcomes, there are limitations to address. Our evaluation was relatively short-term and with a modest sample size. A longer study is needed to confirm that the higher ROI is sustained over months and translates into improved performance in external assessments. There is also the question of how well the gains transfer to language use. We measured words mastered within the platform, but an important educational goal is for learners to actually use these words in writing or speaking. Additionally, while our ROI metric is useful, it treats all words as equal units of learning. In reality, learning a very common, useful word might be more valuable than a rare word. We haven't weighted words by usefulness or difficulty in the ROI. Future research could refine the metric by considering word frequency levels.

 doi:10.20944/preprints202512.2239.v1

From an engineering perspective, the performance of the system was stable during the trial, but at larger scale, we will need to ensure efficiency. Real-time LLM calls can be resource-intensive. Caching strategies might be explored to reduce latency and cost if thousands of users are using the system concurrently. There is ongoing development in distilling large models to smaller ones or using prompt caches for repeated queries. Privacy is another consideration: our current design sends user input to external AI APIs. We mitigated risks by not sending any personally identifying information and by enabling content filters. Techniques like local AI deployment or federated learning could help keep more data on-device in future iterations. Its effectiveness could be amplified by blending with classroom instruction – for instance, a teacher could get analytics from the platform to identify common vocabulary gaps and address them in class. This kind of AI–human collaboration is a promising area.

5. Conclusions

This study presents a novel multimodal AI platform for English vocabulary learning and provides initial evidence of its effectiveness in boosting learning ROI. Learners were able to master vocabulary at an accelerated pace – an outcome attributable to the synergy of adaptive scheduling, LLM-generated content, multimodal practice, and gamified engagement. These results are in line with emerging literature that AI-driven personalized learning can enhance both learning outcomes and learner experiences. Our work pushes the boundary by integrating diverse AI capabilities into one system, highlighting that the whole can be greater than the sum of its parts in educational technology.

Moving forward, we foresee several enhancements and research directions. We plan to conduct long-term studies to measure retention over months and to see if the higher efficiency persists. We are also interested in how such a platform could be used in different contexts – for example, informal learning (outside class) versus formal curriculum integration. A future direction is addressing the human-AI interaction concerns: ensuring transparency of AI recommendations, preventing over-reliance on the AI, and maintaining ethical use. Our current content moderation and the positive results on learner affect suggest that a well-designed AI tutor can indeed be a force for good in education, echoing optimistic views in the field.

In summary, the multimodal personalized vocabulary learning environment introduced here demonstrates a viable path to significantly improving learning ROI in language education. It offers an innovative blend of personalization, multimodal learning, and real-time AI interaction, marking a step towards the next generation of intelligent learning systems. Our findings contribute to the growing body of evidence that AI, when thoughtfully integrated into pedagogical design, can not only accelerate learning outcomes but also enrich the learning experience.

Funding: This work was supported by the National Natural Science Foundation of China (62541330)， the National Natural Science Foundation of China (61272315), the Natural Science Foundation of Zhejiang Province (LY21F020028, LY24F030005), the Science and Technology Project of Zhejiang Province (2023C01040).

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

 doi:10.20944/preprints202512.2239.v1

Appendix A

Appendix A.1

Project documentation list:
App_Data:
es.db
init_es_db.sql
C#RouteConfig.cs
Controllers:
c#AccountController.cs
c#HomeController.cs
ac#ProgressController.cs
c#QuizController.cs
ac# SubscriptionController.cs
Helpers:
C#ChatTtsHelper.cs
c#DbHelper.cs
c#Dblnitializer.cs
c#PracticeData.cs
c#SubscriptionHelper.cs
c#UserHelper.cs
c#VocabQuestionHelper.cs
Models:
ac#Detection.cs
c#PracticeAnswerDetail.cs
c#Question.cs
c#QuestionVm.cs
ac#Result.cs
c#SubjectRankingViewModel.
c#Subscription.cs
c#User.cs
c#UserMastery.cs
c#VocabQuestion.cs
c#WordDetail.cs
c#WordGraphViewModel.cs
Services:
c#AssistantRunService.cs
c#LLMClient.cs
Views:
Account
[@] Login.cshtml
[@]Register.cshtml
Home
[@]Index.cshtml
[@] Ranking.cshtml
[@] WordGraph.cshtml
Progress
[@] Analysis.cshtml
[@]Practice.cshtml
[@]PracticeResult.cshtml
[@]ReportPdf.cshtml

 doi:10.20944/preprints202512.2239.v1

Quiz
[@] Detect.cshtml
[@]Detect.cshtml
[@]Result.cshtml
[@] Sheetlnfo.cshtml
[@] Start.cshtml
Shared
[@]_Layout.cshtml
[@]_ViewStart.cshtml
Subscription
[@]Index.cshtml
Web.config
Global.asax
packages.config
Web.config

References

1. Y. Yang, "AI-supported L2 vocabulary acquisition – a systematic review from 2015 to 2023," *Education and Information Technologies*, **2025,** vol. 30, pp. 17995–18029, DOI: 10.1007/s10639-025-13417-8
2. A. Rahman, A. Raj, P. Tomy, and M. S. Hameed, "A comprehensive bibliometric and content analysis of artificial intelligence in language learning (2017–2023)," *Artificial Intelligence Review*, **2024,** vol. 57, art. 107, DOI: 10.1007/s10462-023-10643-9
3. M. Zhu and C. Wang, "A systematic review of AI in language education: current status and future implications," *Language Learning & Technology*, **2025**, vol. 29, no. 1, pp. 1–29, DOI: 10.64152/10125/73606.
4. E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, **2023**, vol. 8, no. 1. DOI: 10.1016/j.lindif.2023.102274
5. Y. Huang, Z. Zhang, J. Yu, X. Liu, and Y. Huang, "English phrase learning with multimodal input," *Frontiers in Psychology*, **2022**, vol. 13, art. 828022. DOI: 10.3389/fpsyg.2022.828022
6. N. P. Daly, "Investigating learner autonomy and vocabulary learning efficiency with MALL," *Language Learning & Technology*, **2022**, vol. 26, no. 1, pp. 1–30. DOI: 10.64152/10125/73469
7. J. Wang and W. Fan, "The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis," *Humanities and Social Sciences Communications*, **2025,** vol. 12, art. 621. DOI: 10.1057/s41599-025-04787-y
8. J. Meng, "AVAR-RL: adaptive reinforcement learning approach for personalized English vocabulary acquisition," *Discover Artificial Intelligence*, **2025**, vol. 5, art. 317. DOI: 10.1007/s44163-025-00584-3
9. M. H. Al-Khresheh, "The cognitive and motivational benefits of gamification in English language learning: a systematic review," *The Open Psychology Journal*, **2025**, vol. 18, pp. 35–52. DOI: 10.2174/0118743501359379250305083002
10. L. Guan, S. Li, and M. M. Gu, "AI in informal digital English learning: a meta-analysis of its effectiveness on proficiency, motivation, and self-regulation," *Computers and Education: Artificial Intelligence*, **2024**, vol. 7, art. 100323. DOI: 10.1016/j.caeai.2024.100323
11. K. I. Vorobyeva, S. Belous, N. V. Savchenko, L. M. Smirnova, S. A. Nikitina, and S. P. Zhdanov, "Personalized learning through AI: pedagogical approaches and critical insights," *Contemporary Educational Technology*, **2025**, vol. 17, no. 2, art. ep574. DOI: 10.30935/cedtech/16108
12. S. Bhanu, and S. Vijayakumar, "Evaluating AI-personalized learning interventions in distance education," *International Review of Research in Open and Distributed Learning*, **2025**, vol. 26, no. 1, pp. 157–174. DOI: 10.19173/irrodl.v26i1.7813
13. E. Jaleniauskiene, D. Lisaitė, and L. Daniusevičiūtė-Brazaite, "Artificial intelligence in language education: a bibliometric analysis," *Sustainable Multilingualism*, **2023**, vol. 23, DOI: 10.2478/sm-2023-0017

 doi:10.20944/preprints202512.2239.v1

14. J. Yan, C. Wu, X. Tan, and M. Dai, "The influence of AI-driven personalized foreign language learning on college students' mental health: a dynamic interaction among pleasure, anxiety, and self-efficacy," *Frontiers in Public Health*, **2025**, vol. 13, art. 1642608. DOI: 10.3389/fpubh.2025.1642608
15. M. D. Hossain and M. K. Hasan, "A comparative study between the effectiveness of reading-only condition and reading-while-listening condition in incidental vocabulary acquisition," *Journal of Language and Linguistic Studies*, **2022**, vol. 18, Special Issue 1, pp. 277–298.
16. P. Waring and M. Takaki, "At what rate do learners learn and retain new vocabulary from reading a graded reader?," *Reading in a Foreign Language*, **2003**, vol. 15, no. 2, pp. 130–163.
17. H. L. Roediger III and J. D. Karpicke, "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention," *Psychological Science*, **2006**, vol. 17, no. 3, pp. 249–255.
18. N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis," *Psychological Bulletin*, **2006**, vol. 132, no. 3, pp. 354–380.