Computers and Education: Artificial Intelligence 9 (2025) 100505

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Impact of AI-generated storytelling vs. gamified learning on vocabulary retention and engagement in CALL environments

Ehsan Namaziandost [a,*], Fidel Çakmak [b]

[a] Department of English, Islamic Azad University of Ahvaz, Ahvaz, Iran
[b] Department of Foreign Language Education, Alanya Alaaddin Keykubat University, Antalya, Türkiye

ARTICLE INFO

Keywords:
AI-Generated storytelling
Gamified learning
Vocabulary retention
Engagement
CALL environments

ABSTRACT

Artificial Intelligence (AI) and gamified learning have attracted interest from language educators and researchers in the field of teaching English as a Foreign Language (EFL) for their potential to enhance vocabulary acquisition outcomes. However, the relative effectiveness of AI-generated storytelling and gamified learning in EFL vocabulary acquisition has been underexplored. This study examines the influence of AI-generated storytelling and gamified learning on the vocabulary retention and engagement of intermediate-level EFL learners. Ninety participants were allocated among three groups: a control group that employed conventional vocabulary methods (e.g., rote memorization using flashcards and quizzes), a gamified learning group that utilized Duolingo's exercises, and an AI-generated interactive narrative group. Over a four-week period, the gamified group completed interactive exercises, the control group followed conventional instruction, and the AI group engaged with personalized, context-rich narratives that embedded target vocabulary. Data was gathered through pre- and post-tests, which included immediate and delayed vocabulary assessments, a learner engagement questionnaire, and semi-structured interviews. The interviews were analyzed using grounded theory. The AI-generated storytelling group outperformed both the gamified learning and control groups in both immediate and delayed vocabulary tests, as evidenced by quantitative results. This suggests that the AI-generated storytelling group exhibited superior long term vocabulary retention. The control group was also surpassed by the gamified learning group in both assessments, albeit to a lesser extent. The qualitative results suggested that the AI-generated storytelling group reported a higher level of engagement, attributing their motivation to immersive and meaningful narratives. The gamified learning group found the approach to be enjoyable but less profound, and the control group described traditional methods as structured yet uninspiring. These results indicate that AI-generated storytelling is a potent instrument for improving vocabulary acquisition and engagement in EFL settings.

1. Introduction

The prevalence of access to Artificial Intelligence (AI) is reshaping language education practices within the field of English as a Foreign Language (EFL) learning (Bhutoria, 2022; Har, 2023; Namaziandost & Rezai, 2024; Rezai et al., 2024). Among the emerging applications, AI-generated storytelling and gamified learning have attracted considerable attention for their potential to enhance vocabulary acquisition and learner engagement. These methods can provide engaging, interactive experiences that target cognitive understanding, emotional engagement, and active participation, leading to improved learning outcomes. AI-generated narratives provide what Colella (2025) terms 'algorithmic narrativity,' a co-creative practice where human and LLMs shape the story together, and this can reshape the sense of creativity and authorship. This differs from the linear structure of conventional storytelling, allowing learners to actively participate in shaping the narrative cooperatively and, consequently, contributes to their learning journey with positive outcomes.

A primary advantage of AI-generated storytelling is its ability to embed vocabulary within meaningful, narrative contexts, which leads to enhancing comprehension and retention. Traditional vocabulary instruction often relies on rote memorization which is a method that Yang and Dai (2011) frequently criticize for its limited effectiveness in promoting long-term retention. In contrast, AI-generated narratives situate new vocabulary within meaningful, realistic contexts, enabling learners to encounter words and phrases in situations aligned with their interests

* Corresponding author.
E-mail addresses: e.namazi75@yahoo.com (E. Namaziandost), fidel.cakmak@yahoo.com, fidel.cakmak@alanya.edu.tr (F. Çakmak).

https://doi.org/10.1016/j.caeai.2025.100505
Received 13 February 2025; Received in revised form 17 November 2025; Accepted 18 November 2025
Available online 19 November 2025

and learning objectives (Chubb et al., 2022). Furthermore, Cao et al. (2023) argue that the immersive nature of these interactive narratives, combined with personalized feedback, can significantly enhance learner engagement. Supporting this, Wang et al. (2024) found that learners interacting with AI-generated materials report reduced anxiety and greater willingness to communicate in the target language. Additionally, Castro and Shelton-Strong (2024) highlight that emotional connection to the story and the sense of agency afforded by making choices contribute to sustained motivation.

Alongside AI-generated storytelling, gamified learning has emerged as a method for boosting vocabulary retention and engagement. Gamification, which involves integrating game elements into learning activities, capitalizes on the motivational power of games (Nguyen-Viet & Nguyen-Viet, 2025a). This approach exploits the competitive, interactive, and rewarding aspects of games to increase learners' intrinsic motivation to engage with language learning tasks (Nguyen-Viet & Nguyen-Viet, 2025b). In the context of vocabulary acquisition, gamified platforms like Quizlet, Kahoot!, and VocabGame provide dynamic and engaging practice opportunities (Chowdhury et al., 2024). The repetitive exposure to new vocabulary inherent in game-based activities reinforces understanding and strengthens retention (Waluyo & Balazon, 2024).

Moreover, gamification consists of social elements, allowing learners to collaborate or compete with peers (Li et al., 2023). The interaction between peers can build a sense of community and engagement. The adaptability of gamified learning also allows educators to adapt exercises to individual learner needs and proficiency levels (Grabner-Hagen & Kingsley, 2023; Zhou, 2024), providing personalized learning experiences and progress at an appropriate pace. The ability to track progress and receive immediate feedback further contributes to its effectiveness (Dichev & Dicheva, 2017; Kohnke & Moorhouse, 2022).

AI-generated storytelling and gamified learning provide innovative approaches to language education, with demonstrated potential to enhance vocabulary acquisition and learner engagement through personalized and interactive experiences. While AI-generated storytelling provides personalized, context-rich experiences, gamified learning leverages the motivational power of game mechanics. Both approaches have demonstrated positive effects on learner motivation and retention. However, despite these promising developments, researchers have yet to develop a clear understanding of the relative effectiveness of these two approaches within specific EFL contexts. Specifically, the potential for AI-generated storytelling and gamified learning to surpass traditional methods in fostering vocabulary acquisition and sustained engagement requires further investigation. Therefore, this study seeks to investigate the comparative effects of AI-generated storytelling and gamified learning on EFL learners' vocabulary retention and engagement. Understanding these effects is crucial, not only for informing pedagogical practices within EFL classrooms but also for contributing to a broader understanding of the effective integration of AI into foreign language learning environments. To better frame the purpose of this study and guide the exploration of these underexplored areas, the following research questions are posed at this juncture:

1. Do AI-generated storytelling and gamified vocabulary learning significantly improve EFL learners' vocabulary retention?
2. Do AI-generated storytelling and gamified vocabulary learning significantly enhance EFL learners' engagement?
3. How do AI-generated storytelling and gamified vocabulary learning influence EFL learners' vocabulary retention processes?
4. How do AI-generated storytelling and gamified vocabulary learning affect EFL learners' engagement experiences?

2. Literature review

2.1. AI-generated storytelling, vocabulary learning, and engagement

AI-generated storytelling has transformed language education by providing new ways for language learners to enhance vocabulary acquisition and learner engagement. This approach uses AI to create personalized narratives that adapt to individual learners' interests, weaving target vocabulary into meaningful, story-driven contexts (Han & Cai, 2023). According to Chubb et al. (2022), these narratives incorporate adaptive content, multimodal elements like animations and voice recognition, and interactive dialogues that encourage active participation. Such features distinguish AI-generated storytelling from traditional methods, fostering engaging learning experiences that align with students' preferences.

For EFL learners, AI-generated storytelling provides notable benefits in vocabulary learning and engagement. By embedding vocabulary within narrative contexts, this method promotes deeper understanding and retention, as learners can connect words to real-life scenarios rather than simply memorizing them in isolation. Chen (2024) highlights that such contextualized learning improves both immediate recall and long-term retention. This approach also considers the inherent ethical implications of AI use by prioritizing learner-centered design and accessibility, as emphasized by Bender et al. (2021) in their discussion of AI ethics in education. The personalized and interactive nature of AI-driven narratives fosters autonomy and enjoyment, which reduces anxiety and encourages learners to communicate more confidently (Wu & Chen, 2020). Alhamami (2016) note that incorporating visuals and voice interactions further enhances immersion, making vocabulary acquisition more memorable. These features of AI-generated storytelling create a dynamic, learner-focused environment that supports linguistic and motivational growth.

Despite these advancements, research comparing AI-generated storytelling with gamified learning for vocabulary acquisition in EFL settings remains limited. Systematic reviews by Zawacki-Richter et al. (2019) underscore the need for robust empirical studies to evaluate AI's impact on educational outcomes, particularly in vocabulary learning. Studies on digital storytelling, which blends texts, images, and sounds, consistently show its potential to increase engagement and support vocabulary acquisition. For example, Schmoelz (2018) and Hava (2021) found that multimedia narratives enhance interactivity and motivation. Similarly, Kuo-Ping et al. (2018) observed that digital storytelling fosters intrinsic motivation, while Kristiawan et al. (2022) emphasized its effectiveness in teaching academic vocabulary through visual narratives. However, these studies often focus on learner-generated or pre-designed stories rather than AI-driven adaptive narratives. Ramezanzadeh and Ebadi (2025) explored AI-generated storytelling among Iranian EFL learners and noted its role in promoting social-emotional growth, but they did not address vocabulary outcomes. Belda-Medina and Goddard (2024) reported positive attitudes among pre-service teachers toward AI tools in EFL instruction, yet their study also lacked empirical data on vocabulary retention. While these findings highlight the benefits of AI-generated storytelling as regards the affective factors, as discussed by Bender et al. (2021) and Zawacki-Richter et al. (2019), they also present critical ethical perspectives on its systematic integration into education. This study addresses these gaps by examining the effectiveness of AI-generated storytelling, gamified learning, and traditional methods on vocabulary recall, retention, and engagement among EFL learners.

2.2. Gamified learning, vocabulary learning, and engagement

Gamified learning has provided language learners with engaging and interactive experiences. The approach incorporates gaming elements into learning to increase motivation and deepen learner involvement. In EFL settings, features like leaderboards, progress tracking, and

interactive tasks create a dynamic environment that makes vocabulary acquisition more enjoyable and effective (Dichev & Dicheva, 2017; Har, 2022). Research indicates that platforms like Duolingo, while successful in increasing motivation through these mechanics, may fall short in fostering delayed vocabulary retention due to their focus on repetitive, decontextualized exercises (Godwin-Jones, 2019). By turning traditional language instruction into game-like experiences, gamified learning fosters a sense of achievement that encourages students to actively engage with language learning tasks, positively affecting both their motivation and learning outcomes (Jia et al., 2016).

For EFL learners, gamified learning provides clear advantages in vocabulary acquisition and engagement. The use of gaming mechanics, such as earning points or advancing through levels, transforms vocabulary practice into an enjoyable activity, reducing the monotony often linked with memorization. This increased enjoyment fosters intrinsic motivation, encouraging learners to persist in their studies, as Kingsley and Grabner-Hagen (2018) have observed. However, studies indicate that gamified platforms like Duolingo may not fully support delayed retention or meaningful vocabulary use, as they often lack the contextual richness found in narrative-based methods (Loewen et al., 2019). According to Sanosi (2018), Gamified environments also provide instant feedback, which can allow students to observe their learning progress and build confidence, even among those with lower proficiency levels. Furthermore, interactive tasks promote active participation, which strengthens vocabulary recall and nurtures a positive emotional connection to learning (Elaish et al., 2023). Despite these benefits, as Reinhardt (2017) argues, the limited contextual depth in some gamified approaches suggests the need for integrating complementary strategies to enhance vocabulary retention. Overall, gamified learning creates a learner-centered approach that supports linguistic development and sustains engagement in EFL education.

Despite these benefits, prior research on gamified learning in EFL settings reveals mixed outcomes and limitations that underscore the need for further exploration. Several studies have highlighted the potential of digital game-based learning (DGBL) for significantly improved vocabulary gains. For example, Sanosi (2018) and Elaish et al. (2023) found that tools like Quizlet and VocabGame improved vocabulary learning and increased motivation, particularly for learners with lower competence. Setiawan and Wiedarti (2020) reported that Quizlet enhanced classroom engagement, though distractions from mobile phone use posed challenges. Similarly, Sadeghi et al. (2022) observed that gamified instruction increased motivation, but its impact on vocabulary development was modest compared to non-gamified methods. Zhang et al. (2023) researched the patterns of engagement in a DGBVL program, noting positive effects across behavioral, cognitive, and emotional dimensions, but highlighted that these dimensions influenced outcomes differently, indicating complexity in engagement patterns. Rasti-Behbahani and Shahbazi (2022) demonstrated DGBVL's effectiveness in improving productive vocabulary knowledge, yet their focus was on specific word knowledge components rather than broader retention or engagement.

Additional studies further present the benefits and challenges of gamified learning. Smiderle et al. (2020) explored the impact of gamification on engagement, which varied based on students' personality traits, and suggested that individual differences impact the effectiveness of gamified learning. Nadeem et al. (2023) showed that DGBL significantly enhanced engagement and motivation compared to traditional online quizzes, but gender differences and learner variability complicated the results. Anane (2024) reported that Kahoot! increased enthusiasm in foreign language courses, yet the study emphasized general engagement over specific vocabulary outcomes. Waluyo and Balazon (2024) observed that gamified CALL environments improved motivation and enjoyment. However, they noted that competitive elements sustained anxiety, and that psychological factors did not predict linguistic knowledge, raising questions about their impact on vocabulary. While these studies highlight gamified learning's potential to enhance motivation and engagement, they also reveal challenges, including distractions, variable effectiveness across learners, and limited evidence on long term vocabulary retention. Moreover, few studies compare gamified learning with innovative approaches like AI-generated storytelling in EFL contexts. This study addresses the gaps in the field by examining the effectiveness of gamified learning, using Duolingo, in comparison to AI-generated storytelling and traditional methods among EFL learners, with a focus on vocabulary retention and engagement.

Although the preceding points have outlined the distinct contributions of AI-generated storytelling and gamified learning to vocabulary acquisition and engagement in EFL contexts, a synthesis of overlapping themes uncovers common mechanisms that improve contextual learning. This reduces redundancy in the discussion and emphasizes the synergistic potentials for integrated approaches. Both methodologies highlight the significance of integrating vocabulary within meaningful contexts to transcend mere rote memorization. This is illustrated by AI-generated narratives that place words in adaptive, story-driven scenarios (Chen, 2024; Chubb et al., 2022) and by gamified platforms that, though to a lesser degree, include contextual exercises through interactive tasks and feedback (Dichev & Dicheva, 2017; Loewen et al., 2019). The emphasis on contextualization fosters enhanced cognitive processing and facilitates long-term retention, enabling learners to connect new vocabulary with practical applications and emotional experiences. This principle is corroborated by research in both fields (Hava, 2021; Waluyo & Balazon, 2024). Additionally, personalization stands out as a significant theme, with AI storytelling customizing content to align with individual interests and skill levels (Han & Cai, 2023), similar to the adaptive algorithms in gamification that modify difficulty according to performance (Grabner-Hagen & Kingsley, 2023; Zhou, 2024). This approach ultimately enhances intrinsic motivation and alleviates learner anxiety (Elaish et al., 2023; Wu & Chen, 2020). Engagement through interactivity is interconnected, as AI-driven dialogues foster active participation (Schmoelz, 2018) and gamified components such as leaderboards enhance social collaboration (Li et al., 2023). However, it is important to note that gamification frequently emphasizes extrinsic rewards, whereas storytelling utilizes narrative immersion to ensure ongoing involvement. Considerations of ethics, including accessibility and the reduction of bias, create a link between the two, as emphasized in conversations surrounding AI ethics (Bender et al., 2021) and the necessity for inclusive designs in gamification (Smiderle et al., 2020). The literature highlights the potential for hybrid models that merge the contextual richness of AI storytelling with the engaging aspects of gamification. This synthesis could effectively tackle challenges such as the issue of decontextualized repetition found in games and the lack of social components in narratives.

3. Method

3.1. Participants

This study included 90 Iranian EFL learners with B1 level of English proficiency. Participants' ages ranged from 20 to 27 years, and the group consisted of 41 male and 49 female students. To ensure the findings could be universally applied and to account for potential differences in learning environments, participants were selected employing stratified random sampling from three separate language institutes. Riazi (2016) emphasizes the importance of this approach when seeking to represent diverse populations within a study. This sampling method also allowed for consideration of variations in socioeconomic backgrounds, as students were drawn from both public and private institutions, and a range of academic disciplines, including engineering, humanities, and sciences. Before participating, all students completed the Oxford Placement Test to verify their intermediate proficiency in English, aligning with the standards set by the Common European Framework of Reference for Languages (CEFR). Only those scoring within the B1 range were

included, establishing a consistent starting point for vocabulary knowledge across the sample.

Following the confirmation of proficiency, participants were randomly assigned to one of three groups: an AI-generated storytelling group, a gamified learning group, and a control group (CG), with each group containing 30 students. Creswell and Plano Clark (2017) recommend this type of randomization to minimize bias and ensure equal representation across different treatment conditions. Recruitment began with contacting administrators at foreign language institutes to request permission to invite students to participate. Those who expressed interest then received written information about the research, procedures, potential benefits and risks, and students' rights as participants. Informed consent was obtained before the data collection procedure. The form explicitly stated that participation was voluntary and that all data would be kept confidential and anonymized. The participants were also assured that their responses would not affect their grades or standing at the institute.

3.2. Instruments

The first data collection instrument was a pre-test created to evaluate the students' initial vocabulary knowledge prior to the treatment. The pre-test had forty objective items. These items included multiple-choice questions, word-definition matching, and sentence completion tasks, ensuring a comprehensive evaluation of both receptive (word recognition) and productive (word usage) vocabulary knowledge. The pre-test was validated by expert review, ensuring that the items accurately represented the target vocabulary and covered a broad range of usage contexts. Experts in language testing and vocabulary instruction revisited the test for clarity and alignment with the learning objectives. The reliability of the pre-test was established through a pilot test, conducted with a separate sample of fourteen EFL students with B1 level of English. The reliability coefficient, measured utilizing Cronbach's alpha, was .87, confirming strong internal consistency. This implies that the pre-test items were assessing a coherent construct and were suitable for assessing vocabulary knowledge.

The Student Engagement Instrument (SEI) (Appleton et al., 2006) was initially considered for this study because it is a well-established tool for assessing cognitive, emotional, and behavioral engagement in educational settings. It includes 35 items across four subscales (i.e., behavioral engagement, cognitive engagement, emotional engagement, and control and relevance) and uses a 5-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree." Although the SEI has demonstrated strong validity and reliability in general education contexts, the authors recognized that it was not designed specifically for second language learning. As a result, it might not fully capture the distinctive ways in which engagement manifests in EFL vocabulary instruction, such as learners' willingness to interact with narrative input, their investment in acquiring new words, or their responses to digital and gamified tasks. For this reason, the authors chose to use the L2 Engagement Scale, a domain-specific instrument developed to reflect the social, cognitive, and affective dimensions of language learning. This scale was better aligned with the study's focus because it addresses engagement features that are central to language classrooms, including identity, communication, and interaction with language-specific tasks. To ensure the instrument matched the specific aims of the research, the authors adapted it by refining several items to emphasize vocabulary learning and by adding prompts related to AI-generated storytelling and gamified practice. The adapted version preserved the original three-factor structure, behavioral, cognitive, and emotional engagement, but contextualized each item within the EFL vocabulary setting. Before finalizing the instrument, the authors piloted it with a group of 20 Iranian EFL learners who were not part of the main study. Based on their feedback, minor adjustments were made to improve clarity and cultural appropriateness. The final version demonstrated strong internal consistency, with a Cronbach's alpha of .84.

The third instrument in this study was a vocabulary post-test that assessed students' vocabulary retention after the treatment. This post-test included the target words that were introduced to the participants during the four-week treatment period and was employed as both an immediate and a delayed post-test. The immediate post-test was given right after the treatment phase, while the delayed post-test was administered eight weeks later. The post-test consisted of forty items developed to assess both receptive and productive vocabulary knowledge. These items included word-definition matching, where students matched each vocabulary word with its corresponding definition, and sentence completion tasks, where learners were required to use the vocabulary words in context. There were also multiple-choice questions to assess the learners' ability to recognize the correct meaning or usage of the words. Subject-matter experts carefully reviewed the post-test to ensure its validity, confirming that the test items accurately reflected the vocabulary taught during the treatment. The post-test reliability coefficient was .89 as computed using Cronbach's alpha.

In addition to the pre- and post-tests, interviews were conducted with a total of fifteen participants from three groups to explore their experiences and views about the vocabulary learning interventions. The interviews aimed to reveal how students engaged with the interventions, which elements of AI-storytelling and gamified learning they found most effective, and how these experiences affected their vocabulary learning, engagement in tasks and motivation for learning practices. The interviews were designed to permit participants to express their thoughts in depth, offering more detailed insights into their learning experiences that quantitative assessments could not provide.

Each interview lasted between 17 and 25 min and were conducted until data saturation was reached after interviewing a subset of learners from each group. Data saturation was met once the responses from each group became repetitive and no new insights emerged. The interviews were recorded digitally and transcribed. The transcriptions were then coded and analyzed to determine the main themes and patterns that emerged from the participants' responses. The interview questions were structured to include open-ended prompts designed to address key areas of interest linked to the interventions. These questions were validated through expert reviews to guarantee their clarity and alignment with the study's objectives. Questions incorporated expert feedback to ensure that they efficiently captured learners' perceptions and experiences. This validation process ensured the reliability and accuracy of the qualitative data collected.

3.3. Data collection procedures

This study implemented a structured process to collect data and ensure consistent delivery of three distinct interventions for EFL learners. To maintain intervention fidelity, instructors for all groups received thorough training prior to the study. This training focused on standardized delivery and accurate task implementation, ensuring uniformity across the groups. Regular platform consistency checks were also conducted to confirm that both the AI-generated storytelling system and Duolingo functioned as intended and delivered content reliably. A pre-test established participants' baseline vocabulary knowledge. Participants, selected for their intermediate English proficiency based on a standardized placement test, were then randomly assigned to one of three groups: AI-generated storytelling, gamified vocabulary learning via Duolingo, or a CG.

The AI-generated storytelling group participated in a personalized vocabulary learning experience using an AI-powered platform over four weeks. The system was designed utilizing a substantial language model (LLM) akin to GPT-4, employing transformer-based frameworks for the creation of narratives. The algorithms utilized attention mechanisms and autoregressive decoding to create coherent narratives: input prompts integrated target vocabulary lists with narrative frameworks, generating text token-by-token while ensuring contextual coherence. The system employed user-input parameters for personalization

mechanics, including chosen themes (e.g., travel, family), proficiency level (B1 CEFR), and learning objectives to dynamically modify prompts. This process included few-shot prompting methods, where sample narratives were provided to the model to steer the output, along with reinforcement learning from human feedback (RLHF) to enhance relevance and engagement. Vocabulary embedding was accomplished by incorporating words into essential plot points, utilizing adaptive difficulty scaling informed by user performance logs (e.g., correct or incorrect responses prompting the use of simpler or more complex sentences). To ensure reproducibility, prompts adhered to a specific structure: "Create an interactive story centered on [theme] at B1 level, incorporating the following words: [vocabulary list]. Incorporate options that allow the user to shape the narrative. The open-source characteristics of comparable models (e.g., through the Hugging Face Transformers library) enable researchers to reproduce this configuration with limited adjustments. They acquired 100 new vocabulary terms, presented at a pace of 25 each week. The platform developed interactive narratives that integrated target vocabulary into captivating storylines, enabling learners to shape the plot's direction through their decisions. This contextual approach sought to enhance comprehension and retention, aligning with Mayer's (2005) principles of multimedia learning. The AI platform offered prompt corrective feedback, proposed different phrasing options, and monitored individual progress to create customized review exercises seamlessly integrated into the narrative, enhancing difficult vocabulary and aiding in long-term retention.

Alongside these features, the study included various measures to tackle ethical considerations associated with AI-driven educational tools, with a particular emphasis on algorithmic bias and data privacy, both of which are essential for upholding fairness and trust in language learning settings. To address potential algorithmic bias, the AI platform was chosen following an evaluation of its training data sources to guarantee diverse linguistic representations, preventing an overreliance on prevailing English dialects that might disadvantage nonnative EFL learners from different cultural backgrounds; this action is consistent with ethical AI deployment guidelines in education, highlighting the importance of inclusivity and equity in content creation. Additionally, the research team performed regular evaluations of the produced narratives to identify and rectify any unintended biases in story themes or vocabulary integration, utilizing recognized frameworks that promote transparent AI practices to avert discriminatory results in educational resources. Concerning data privacy, all interactions with the AI platform adhered to stringent protocols that aligned with international standards, including the General Data Protection Regulation (GDPR). This ensured that learner inputs, progress data, and personal preferences were encrypted during transmission and securely stored on servers, with access restricted to authorized researchers only. Participants received comprehensive information regarding data usage via detailed consent forms, which specified how their anonymized contributions could aid in enhancing AI models. Additionally, they were given the option to request data deletion, ensuring their autonomy and confidentiality were respected. The implementation of these ethical safeguards served to protect participants while simultaneously bolstering the study's credibility. This alignment with current standards in AI ethics promotes a responsible integration of technology within EFL instruction.

As learners progressed, the AI platform dynamically adjusted story complexity. Once vocabulary was mastered, narratives introduced more demanding scenarios, such as dialogues that required precise lexical choices, to further stretch learners' abilities. This adaptive scaffolding maintained engagement while accommodating individual learning paces, fostering both vocabulary growth and critical thinking. To enhance motivation, the intervention incorporated light gamified elements: learners earned "story points" for completing narrative arcs or making accurate word choices, which could be exchanged for customization options like character avatars or thematic settings. Weekly progress reports highlighted vocabulary gains and narrative milestones, offering a sense of achievement and continuity.

To better understand which specific features contributed most to the observed learning gains, the research team examined platform usage logs alongside qualitative data from post-intervention interviews. The logs showed that participants spent more time on story segments where target words appeared in meaningful, emotionally resonant contexts, and they revisited difficult words more often when those words were tied to key decision points in the narrative. In interviews, learners frequently explained that they remembered words because of what happened in the story or what their character had to say. These responses suggest that the narrative context, particularly its coherence, emotional relevance, and episodic structure, served as a powerful cognitive anchor for vocabulary retention. At the same time, learners expressed appreciation for the ability to shape story outcomes, indicating that interactivity increased their sense of ownership and attention. Personalization also played a role: during onboarding, learners selected themes of interest such as travel, university life, or social relationships, and the AI generated stories aligned with those preferences. This alignment appeared to heighten relevance and motivation. While narrative context, interactivity, and personalization were closely intertwined in the design, the evidence points to narrative context as the foundational element, with interactivity and personalization acting as supportive mechanisms that amplified engagement and cognitive processing.

The gamified learning group used Duolingo, a platform known for its game-based approach to language learning. Duolingo was selected over alternatives like Quizlet or Kahoot because of its adaptive algorithms, diverse exercise types, and motivational features such as points, streaks, and leaderboards, all of which aligned with the study's focus on gamified instruction. Participants learned the same 100 target vocabulary words over four weeks, at a pace of 25 words per week. The platform's varied activities, including matching, sentence completion, and multiple-choice quizzes, provided repeated exposure in different formats, which likely supported retention through distributed practice. Duolingo's algorithm adjusted task difficulty based on performance, gradually introducing more complex exercises such as open-ended sentence construction. Features like streaks and leaderboards encouraged daily engagement and fostered a sense of friendly competition. In-app explanations of word meanings and usage tips further enriched the learning experience. However, it is worth noting that Duolingo occasionally includes short animated stories as part of its lessons. These narrative snippets, though minimal, introduced an unintended narrative dimension that may have partially overlapped with the storytelling condition.

CG adhered to conventional vocabulary acquisition techniques, employing a compilation of precisely one hundred target words, each paired with a distinct definition and an illustrative example sentence. Consistent with recognized methodologies in EFL teaching that prioritize systematic repetition for developing core knowledge, research has highlighted the use of paper-based resources for vocabulary learning. This approach focused on flashcards and weekly assessments, emphasizing repetition and memorization over engagement or individualization. These activities were exclusively conducted on paper and facilitated through instruction led by the teacher in the classroom setting. At the beginning of every week, the instructor distributed a printed handout featuring 25 new vocabulary terms to the students. Every term was accompanied by its definition and an illustrative example sentence sourced from standard academic English. To manage time efficiently and guarantee regular exposure, the weekly sessions followed a set timetable that reflected typical traditional EFL practices, where the initial phases concentrated on introduction and encoding. In the Monday session, which spanned 45 min to ensure comprehensive preparation, students engaged in the hands-on creation of physical flashcards: they inscribed the target word on one side and transcribed the definition along with an example sentence on the reverse side. The instructor oversaw this procedure to guarantee precision and uniformity

among students.

On Wednesdays, students engaged in a 15-min session where they partnered up to quiz one another with their flashcards. This pair work exercise, a fundamental component of traditional vocabulary drills that encourages collaborative recall and instant feedback, required students to alternate roles. One student would present the word side while the partner would offer the definition or use it in a sentence, followed by a discussion of any mistakes made. The instructor circulated throughout the classroom to monitor interactions and provide timely corrections as necessary. Subsequently, a brief quiz was conducted during class on Fridays. The assessment was structured to evaluate cumulative retention using diverse formats that mirror conventional evaluation methods. Each quiz was timed at 10 min and comprised three types of tasks: matching words to their definitions, filling in sentences with the appropriate target word, and crafting an original sentence that incorporated a specified word. The matching task involved students pairing 10 words with their definitions from a mixed list, while the sentence completion required them to fill in blanks in 10 given sentences to assess contextual understanding. Additionally, the original sentence writing encouraged learners to create 5 sentences showcasing productive use, all within a set timeframe to replicate the real-time recall pressure often experienced in traditional EFL environments. The instructor gathered, evaluated, and handed back the quizzes in the next class, accompanied by concise written feedback. The absence of a digital component, lack of adaptive feedback, and no chance for self-paced review beyond the classroom were notable shortcomings. This organized, instructor-focused method, despite its absence of the dynamic features present in the experimental groups, established a foundation for comparison by utilizing established repetitive techniques that have demonstrated effectiveness in facilitating basic memorization in EFL settings, albeit frequently at the cost of more profound engagement.

3.4. Data analysis

This study used quantitative and qualitative data analysis. For the quantitative data, the descriptive statistics were calculated to characterize the participant sample and provide a general overview of performance across all three groups on the pre- and post-tests, as well as the engagement questionnaire. Subsequently, authors applied inferential statistics to the findings. For this purpose, they conducted a one-way Analysis of Variance, or ANOVA. All statistical analyses were conducted using SPSS version 26. Before proceeding with the one-way ANOVA, the underlying assumptions of normality, homogeneity of variance, and homogeneity of regression slopes were carefully examined to ensure the validity of the results. Tests for normality, using the Shapiro-Wilk statistic, and homogeneity of variance, using Levene's test, revealed no substantial violations of these assumptions, with p-values exceeding .05. Furthermore, the absence of significant interaction terms indicated that the regression slopes were sufficiently similar across the groups. Authors then ran the one-way ANOVA.

To complement the quantitative findings, the qualitative data from semi-structured interviews were analyzed adopting a grounded theory approach (Strauss & Corbin, 1990). This method was opted for its potential to identify emergent themes, and to carry out an in-depth exploration of participants' views on AI-generated storytelling, gamified learning, and traditional instructional methods. The analysis started with open coding, where the transcripts were read and segmented into meaningful units, each assigned descriptive codes like "engagement with AI-generated storytelling" or "motivation in gamified tasks," derived from the data. During axial coding, these codes were grouped into broader categories. In the selective coding phase, categories were sorted into a core category, such as "transformative learning through AI and gamification," which was refined to address the research questions and draw meaningful inferences. To ensure reliability, 25 % of the transcripts were independently coded, achieving a Cohen's Kappa coefficient of .78, indicating substantial agreement. Validity was confirmed through member checking, with 9 participants confirming the accuracy of themes and quotations.

4. Results

This section presents the quantitative and qualitative findings of the study.

4.1. Quantitative results

To establish baseline equivalence across the three groups, descriptive statistics for the pretest vocabulary scores were calculated and are presented in Table 1.

As shown in Table 1, the pretest mean scores for vocabulary knowledge were closely aligned across the groups, with the AI-generated storytelling group recording the highest mean (M = 10.25, SD = 1.73), followed by the gamified group (M = 10.10, SD = 2.07), and the control group (M = 9.85, SD = 2.21). The standard deviations indicate comparable variability in scores within each group, suggesting that the groups were homogeneous in their initial vocabulary knowledge. To confirm the statistical equivalence of the groups at baseline, a one-way ANOVA was run on the pretest scores, with results presented in Table 2.

The ANOVA results in Table 2 indicate no statistically significant differences in pretest vocabulary scores among the three groups, $F(2, 87) = .301$, $p = .74$, partial $\eta^2 = .006$. The non-significant p-value ($p > .05$) and small effect size (partial $\eta^2 = .006$) confirm that the groups had comparable vocabulary knowledge prior to the intervention, ensuring a fair basis for evaluating the effects of the treatments. To evaluate the effectiveness of the interventions on vocabulary retention, an immediate post-test was administered following the four-week treatment period. Descriptive statistics for the retention assessment are presented in Table 3.

As shown in Table 3, the AI-generated storytelling group achieved the highest mean score on the immediate post-test (M = 16.80, SD = 1.73), followed by the gamified group (M = 15.30, SD = 2.05), and the control group with the lowest score (M = 12.00, SD = 1.70). The relatively small standard deviations across groups suggest consistent performance within each condition, with slightly greater variability in the control group. These results indicate that the AI-generated storytelling intervention may have been particularly effective in fostering immediate vocabulary retention, followed by the gamified learning approach, while the traditional methods used by the control group yielded the least improvement. To determine whether these differences were statistically significant, a one-way ANOVA was conducted, with results reported in Table 4.

The ANOVA results in Table 4 reveal a statistically significant effect of the intervention type on vocabulary retention, $F(2, 87) = 53.69$, $p < .00$, partial $\eta^2 = .549$. The large effect size (partial $\eta^2 = .549$) indicates that approximately 54.9 % of the variance in retention scores can be attributed to the intervention type, underscoring the substantial impact of the treatments. This significant result suggests that at least one group performed differently from the others, prompting further analysis to identify specific group differences. A post hoc Scheffé test was conducted to explore pairwise comparisons among the groups, with results presented in Table 5.

As indicated in Table 5, the post hoc Scheffé test confirmed significant differences across all pairwise comparisons. The AI-generated storytelling group outperformed the gamified learning group (mean

Table 1
Descriptive statistics for vocabulary knowledge (pretest).

Group	N	Mean	SD
AI-generated storytelling	30	10.25	1.73
Gamified	30	10.10	2.07
Control	30	9.85	2.21

Table 2
One-way ANOVA for vocabulary knowledge (pretest).

Source	df	F	p	Partial η^2
Between Groups	2	.301	.74	.006
Within Groups	87			

Table 3
Descriptive statistics for immediate posttest (retention assessment).

Group	N	Mean	SD
AI-generated storytelling	30	16.80	1.73
Gamified	30	15.30	2.05
Control	30	12.00	1.70

Table 4
One-way ANOVA for immediate posttest (retention assessment).

Source	df	F	p	Partial η^2
Between Groups	2	53.69	<.00	.549
Within Groups	87			

Table 5
Post hoc Scheffé test for immediate posttest.

Group Comparison	Mean Difference	p	Cohen's d	Effect Size Description
AI-generated storytelling vs. Gamified	1.50	<.05	.79	Moderate
AI-generated storytelling vs. Control	4.80	<.05	2.80	Very large
Gamified vs. Control	3.30	<.05	1.75	Large

difference = 1.50, $p < .05$) and the control group (mean difference = 4.80, $p < .05$), with the latter comparison reflecting a particularly significant difference. Additionally, the gamified learning group significantly surpassed the control group (mean difference = 3.30, $p < .05$). To further quantify the magnitude of these differences, Cohen's d was calculated for each comparison: AI-generated storytelling vs. gamified learning (d = .79, moderate effect), AI-generated storytelling vs. control group (d = 2.80, very large effect), and gamified learning vs. control group (d = 1.75, large effect). These findings confirm that AI-generated storytelling was the most effective intervention for enhancing immediate vocabulary retention, followed by gamified learning, both of which outperformed the traditional methods used by the control group. To assess the effects of the interventions on long term vocabulary retention, a delayed post-test was administered eight weeks after the treatment period. Descriptive statistics for the retention assessment are presented in Table 6.

As shown in Table 6, the AI-generated storytelling group recorded the highest mean score on the delayed post-test (M = 15.90, SD = 2.15), followed by the gamified learning group (M = 14.20, SD = 2.12), and the control group with the lowest score (M = 10.90, SD = 2.63). The standard deviations indicate comparable variability across groups, with slightly greater spread in the control group. These findings indicate that AI-generated storytelling significantly improved long term vocabulary retention compared to gamified learning and traditional methods, with the gamified learning group also outperforming the control group. To determine whether these differences were statistically significant, a one-way ANOVA was conducted, with results reported in Table 7.

Table 6
Descriptive statistics for delayed posttest (retention assessment).

Group	N	Mean	SD
AI-generated storytelling	30	15.90	2.15
Gamified	30	14.20	2.12
Control	30	10.90	2.63

The one-way ANOVA results in Table 7 indicate a statistically significant effect of intervention type on vocabulary retention, $F(2, 87) = 36.185$, $p < .05$, partial $\eta^2 = .454$. The large effect size (partial $\eta^2 = .454$) suggests that approximately 45.4 % of the variance in retention scores is attributable to the intervention type, highlighting the substantial impact of the treatment. For further analysis to identify specific group differences, a post hoc Scheffé test was performed to explore pairwise comparisons among the groups, with the results presented in Table 8.

As indicated in Table 8, the post hoc Scheffé test confirmed significant differences across all pairwise comparisons. The AI-generated storytelling group outperformed the gamified learning group (mean difference = 1.70, p = .021) and the control group (mean difference = 5.00, $p < .05$), with the latter comparison showing a particularly pronounced difference. The gamified learning group also significantly surpassed the control group (mean difference = 3.30, p = .000). To quantify the magnitude of these differences, Cohen's d was calculated: AI-generated storytelling vs. gamified learning (d = .80, moderate effect), AI-generated storytelling vs. control group (d = 2.08, very large effect), and gamified learning vs. control group (d = 1.38, large effect). These findings underscore that AI-generated storytelling was the most effective intervention for sustaining long term vocabulary retention, followed by gamified learning, both of which outperformed the traditional methods used by the control group. To assess learner engagement following the four-week intervention, participants completed a questionnaire measuring engagement on a 0–100 scale. Descriptive statistics for the engagement scores are presented in Table 9.

As shown in Table 9, the AI-generated storytelling group achieved the highest mean engagement score (M = 85.50, SD = 4.95), followed by the gamified learning group (M = 77.00, SD = 4.47), and the control group with the lowest score (M = 57.90, SD = 5.01). The standard deviations indicate moderate variability within each group, with the control group showing slightly greater spread. These results suggest that AI-generated storytelling was the most engaging intervention, due to its personalized and narrative-driven approach. The gamified learning group also demonstrated notable engagement, while the control group, relying on traditional methods, reported the lowest levels of engagement. To examine whether these differences were statistically significant, a one-way ANOVA was conducted on the engagement scores. The results are presented in Table 10.

According to the ANOVA results in Table 10, the type of intervention had a statistically significant effect on engagement ($F(2, 87) = 248.777$, $p < .05$, partial $\eta^2 = .851$). The large effect size (partial $\eta^2 = .851$) of the treatments demonstrates their substantial impact, indicating that the type of intervention accounts for approximately 85.1 % of the variance in engagement scores. To examine specific group differences (hypothetical, as indicated by the reported significant differences), a post hoc Scheffé test was implemented. Significant pairwise differences were observed between AI-generated storytelling group and gamified learning group (mean difference = 8.50, $p < .05$, d = 1.75, large effect), AI-generated storytelling and control group (mean difference = 27.60, $p < .005$, d = 5.54, extremely large effect), and gamified learning group and control group (mean difference = 19.10, $p < .05$, d = 3.92, very large effect). These findings indicate that AI-generated storytelling and gamified learning were the most engaging interventions.

Table 7
One-way ANOVA for delayed posttest (retention assessment).

Source	df	F	p	Partial η^2
Between Groups	2	36.185	<.05	.454
Within Groups	87			

Table 8
Post hoc Scheffé test for delayed posttest.

Group Comparison	Mean Difference	p	Cohen's d	Effect Size Description
AI-generated storytelling vs. Gamified	1.70	.021	.80	Moderate
AI-generated storytelling vs. Control	5.00	<.05	2.08	Very large
Gamified vs. Control	3.30	.000	1.38	Large

Table 9
Results of engagement analysis.

Group	N	Mean	SD
AI-generated storytelling	30	85.50	4.95
Gamified	30	77.00	4.47
Control	30	57.90	5.01

Table 10
One-way ANOVA for engagement scores.

Source	df	F	p	Partial η^2
Between Groups	2	248.777	<.05	.851
Within Groups	87			

4.2. Qualitative results

The AI-generated storytelling group demonstrated overwhelmingly positive responses to the intervention, consistently highlighting its immersive, engaging, and context-rich nature. Participants described the experience as feeling like an adventure, where vocabulary was naturally embedded within compelling narratives, making words easier to understand and remember. As one participant noted, "*The tales seemed like adventures. The words were easier to remember because I could see them in the story*," illustrating how narrative context strengthened cognitive associations and supported long term retention. Others emphasized the emotional resonance of personalized content, with one stating, "*I was excited for each session because the stories were about things I like, like family and travel.*" This personal relevance not only deepened engagement but also aligned with the group's high quantitative engagement scores, suggesting that emotionally meaningful content played a key role in sustaining motivation.

Further insights revealed the intervention's effectiveness in promoting real-world language use and reducing the perceived burden of learning. Participants reported being able to visualize scenes from the stories, which helped them recall and use vocabulary in conversations, demonstrating successful transfer beyond the learning environment. One learner shared, "*When I read a word in the story, I could picture the scene, which helped me use it later when I talked to people*," underscoring the cognitive scaffolding provided by vivid storytelling. The game-like, enjoyable format also transformed the learning experience, "*It didn't seem like studying. It was fun, like reading a book or playing a game, but I think I learned a lot*," indicating that reduced cognitive load and increased enjoyment contributed to sustained participation. Finally, the adaptive nature of the AI, which tailored stories to individual interests, was credited with boosting attention and memory, as expressed in the statement, "*The AI made the stories fit what I liked, which made me pay more attention and remember the words better*." Together, these qualitative insights confirm that the AI-generated storytelling intervention succeeded by blending personalization, narrative immersion, and interactivity, resulting in superior vocabulary retention and engagement.

Participants in the gamified learning group, who used Duolingo's vocabulary exercises, reported positive experiences characterized by interactivity and motivation. However, their engagement was less intense compared to those in the AI- generated storytelling group. Many appreciated the app's game-like features, such as earning points and progressing through levels, which created a sense of achievement and encouraged continued use. As one participant noted, "*I liked the app because I could see how far I had come and earn points*," highlighting how gamification fostered intrinsic motivation and contributed to moderate engagement levels. Another shared, "*I enjoyed pushing myself to reach the next level. It made me want to keep going*," illustrating how self-directed challenges supported sustained participation. These motivational elements align with the group's improved performance over the CG in the quantitative results, suggesting that the interactive design effectively promoted initial learning and engagement.

Despite these benefits, participants also expressed limitations related to the lack of meaningful context in vocabulary practice. Several noted that while the exercises were quick and enjoyable, the words often failed to stick. One participant observed, "*It was easy and quick to do the exercises, but I didn't always know how to use the words outside of the app*," pointing to a gap between isolated practice and real-world application. Another remarked, "*It seemed like a game, but some words didn't stick because they weren't part of a bigger story or idea*," emphasizing the absence of narrative depth that could reinforce deeper cognitive processing. This lack of contextual embedding likely hindered long term retention and transfer, explaining the group's lower retention scores compared to the AI-generated storytelling group. Overall, while the gamified learning approach successfully enhanced motivation and accessibility, its limited contextual richness constrained the depth and durability of vocabulary learning.

The CG reported the least positive experiences, with participants consistently describing the process as monotonous, demotivating, and disconnected from real-life use. One participant noted, "*I got bored quickly because all I had to do was memorize lists of words*," reflecting the low engagement and poor retention that align with the group's lowest quantitative scores. Another shared, "*I tried to learn the words, but it was hard to stay interested without a story or game*," highlighting how the absence of interactive or contextual elements diminished motivation and hindered sustained effort. A third participant observed, "*The exercises seemed to be the same over and over, and I couldn't see how the words could be used in real life*," underscoring the lack of meaningful context that limited both engagement and the ability to transfer vocabulary to authentic situations. Collectively, these responses reveal that traditional methods were perceived as passive and decontextualized, offering little cognitive or emotional stimulation, which helps explain the group's inferior performance in retention and overall engagement compared to both the AI-generated storytelling and gamified learning groups.

5. Discussion

This study investigated whether AI-generated storytelling and gamified learning could improve vocabulary retention and engagement among EFL learners compared to traditional methods. The results showed that the AI-generated storytelling group significantly outperformed both the gamified learning group and the CG in immediate and delayed post-tests for vocabulary retention as well as in student engagement. The gamified learning group also outperformed the CG in these areas, though its impact was less marked. Qualitative interviews revealed that learners in the AI-generated storytelling group found the approach highly motivating and immersive. In contrast, those in the gamified learning group described their experience as enjoyable but superficial, and the CG viewed traditional methods as structured but less engaging. These findings indicate that AI-generated storytelling, with its rich and interactive narratives, offers a highly effective and engaging method for vocabulary acquisition. At the same time, gamified learning approaches provide moderate improvements over conventional instruction. Nonetheless, a significant constraint emerges from the gamified condition's implementation of Duolingo, which includes brief

animated narratives as additional components. The brief narrative fragments, though limited, present a possible confounding factor due to their partial overlap with the storytelling method generated by AI. This could have blurred the lines between the two experimental conditions, possibly leading to inflated retention and engagement scores for the gamified group as a result of unintended narrative exposure. While the Duolingo stories are concise and not the main emphasis of its gamified activities, their inclusion might clarify why the advancements in the gamified group, although notable compared to the control, were not as marked as those in the purely narrative-focused AI condition. This perplexing element highlights the necessity for meticulous platform choice in comparative research to differentiate gamification impacts from narrative effects.

The impressive performance of the AI-generated storytelling group aligns with several studies on AI-enhanced and narrative-based language learning. For example, Cho and Yim (2024) found that AI-generated stories significantly improved preschool children's vocabulary skills, a result that mirrors the enhanced vocabulary retention and engagement in this study. Belda-Medina and Goddard (2024) reported that AI-generated storytelling creates engaging and relevant materials for EFL learners, supporting the current evidence that AI-generated storytelling increases engagement. Similarly, Koraishi (2024) noted that ChatGPT-assisted narrative writing improved writing fluency and overall performance in EFL contexts, aligning with the improved vocabulary retention in the AI-generated storytelling group. Samsul et al. (2023) also highlighted the role of AI-driven learning analytics in optimizing educational outcomes, which complements the personalized and data-driven nature of AI-generated storytelling in this study.

Additional research provides further support. Nami and Asadnia (2024) showed that self-made digital stories enhanced EFL learners' vocabulary acquisition, paralleling the benefits of AI-generated storytelling demonstrated in this study, as both approaches use narrative contexts to support learning. Chen (2024) also found that AI-generated storytelling applications positively impacted language acquisition in early childhood education. Similarly, Beguš (2024) likewise demonstrated the effectiveness of AI-generated storytelling in foreign language learning; their findings were that AI-generated storytelling significantly improved vocabulary acquisition and learner engagement. Fu et al. (2022) noted that digital storytelling improved EFL learners' speaking competence, supporting the broader benefits of narrative-based approaches, as evidenced by qualitative feedback in this study, where learners reported high motivation and immersion with AI-generated storytelling.

The effectiveness of AI-generated story-telling aligns with the principles of Cognitive Load Theory. This theory posits that learning is enhanced when cognitive demands are reduced by presenting vocabulary within well-structured and contextually meaningful narratives (Sweller, 1988). AI-generated stories provide vocabulary within coherent and contextually rich narratives, reducing unnecessary cognitive load and allowing learners to focus on understanding and retaining unfamiliar words, as demonstrated by improved retention outcomes. The clear organization of AI-generated narratives helps to lessen unnecessary cognitive strain by eliminating irrelevant details and arranging the content in a logical manner. This organization allows learners to allocate more of their working memory to activities that are beneficial for learning, like building schemas for integrating vocabulary. Narrative coherence is crucial in reducing cognitive load as it creates predictable patterns and logical progressions within the story. This enables learners to anticipate events and integrate new vocabulary more smoothly, minimizing the mental effort required to navigate disjointed elements. When narratives are coherent, they help learners process information more easily by fitting into their existing knowledge frameworks. This alignment reduces the mental effort needed to understand complex language, ultimately improving comprehension. The connection is especially evident in the improved retention outcomes observed in the study. Participants showed a decrease in cognitive load while learning new vocabulary, attributed to the seamless integration of words into meaningful narratives that fostered focused attention and deeper comprehension. Furthermore, recent studies on AI-driven tools enhance the application of Cognitive Load Theory, emphasizing the importance of adaptive narratives in efficiently managing cognitive resources during language tasks (Fan & Yao, 2025). For example, in environments where AI assists with language learning, tools designed to maintain narrative coherence can alleviate some of the cognitive load by offering structured prompts and assessments. This allows learners to concentrate on more complex tasks, such as synthesizing and applying vocabulary effectively. The relationship is further strengthened by frameworks that focus on coherence in AI narratives, as ensuring story consistency plays a crucial role in minimizing cognitive strain and enhancing learning efficiency (Yi et al., 2025). Expanding upon this foundation, Constructivist Learning Theory aligns with these findings by highlighting the importance of actively constructing knowledge through engagement with meaningful content (Piaget, 1970). The storytelling approach generated by AI invites learners to actively participate in personalized narratives, promoting a deeper understanding of the material. This is supported by qualitative feedback from participants, who describe the method as both motivating and immersive. Abuhassna et al. (2021) observed that learning environments enhanced by technology, particularly those incorporating interactive modules, foster engagement by offering structured and dynamic experiences that resonate with the immersive qualities of AI-generated storytelling.

Multimedia Learning Theory further explains the study's results, asserting that combining verbal and visual elements enhances learning outcomes (Mayer, 2005). AI-generated storytelling integrates text, audio, and visuals, creating a rich, immersive experience that supports stronger vocabulary encoding and higher engagement, as evidenced by learners' positive perceptions reported in interviews. Motivational Theory also plays a role, as the interactive and narrative-driven nature of AI-generated storytelling sustains learner interest, contributing to elevated engagement levels (Dörnyei & Ushioda, 2021). Narrative Theory highlights the power of stories to make information relatable and memorable (Bruner, 1991). By embedding vocabulary in engaging narratives, AI-generated storytelling helps learners connect words to real-life contexts, which accounts for the superior vocabulary retention and positive learner attitudes.

The study also found that the gamified learning group outperformed the CG in vocabulary retention and engagement on both immediate and delayed post-tests among EFL learners. These results align with several studies exploring gamification's role in language education. For example, Saleem et al. (2022) observed that gamified instruction significantly improved vocabulary learning and motivation, resulting in the enhanced vocabulary gains and engagement seen in this study's gamified learning group. Huang et al. (2020) also found that gamified approaches increased cognitive outcomes, such as vocabulary acquisition, and motivational outcomes, like engagement, which supports the current findings. Similarly, Waluyo and Bucol (2021) reported that gamified instruction improved vocabulary gains, providing an effective learning environment for EFL learners.

Additional research strengthens these findings. Zainuddin et al. (2024) noted that gamified courses boosted student motivation compared to traditional classrooms, aligning with this study's qualitative data where learners described the gamified approach as enjoyable, though superficial. Landers et al. (2017) found that gamified elements enhanced engagement and performance in language learning, consistent with the higher engagement levels reported by the gamified learning group. Sadeghi et al. (2022) also observed that gamified instruction improved vocabulary gains and motivation, while Li and Jianxing (2025) showed that gamified learning systems enhanced learning behavior and achievement. Fotaris et al. (2016) highlighted competition and fun as key motivators in gamified learning settings, a sentiment echoed in the qualitative feedback from this study's gamified learning

group, which emphasized enjoyment as a driving factor.

The success of gamified learning can be explained through the Attention, Relevance, Confidence, and Satisfaction (ARCS) model of motivation, which emphasizes positive reinforcement to sustain learner engagement (Keller, 1979). The gamified approach provided immediate feedback through rewards like badges, consistent with Reinforcement Theory, which posits that positive reinforcement encourages repeated engagement with vocabulary tasks (Skinner, 1965). This might have contributed to the increased engagement and vocabulary retention observed in the gamified learning group. Self-Determination Theory further supports these results, as gamified systems foster autonomy and competence through elements like progress tracking, which enhance learners' motivation (Ryan & Deci, 2024; 2020). The qualitative feedback, where learners noted enjoyment, reflects how these motivational elements encouraged sustained participation and vocabulary acquisition.

Social Cognitive Theory also provides insight, suggesting that social interactions and comparisons drive performance (Bandura, 2001). The gamified learning group's improved outcomes likely stemmed from competitive elements, such as leaderboards, which motivated learners to outperform peers, as outlined by the theory of social comparisons (Festinger, 1954). Constructivist Learning Theory and Flow Theory further explain the findings, as gamified systems adjust task difficulty to match learners' abilities, creating an engaging and balanced challenge that supports vocabulary acquisition (Csikszentmihalyi & Larson, 2014; Vygotsky, 1978). Due to its adaptive content and guidance, gamification aligns with scaffolding principles, marking learners progress within their zone of proximal development, which accounted for the observed improvements in vocabulary retention.

6. Conclusions and implications

The findings indicate that AI-generated storytelling is the most effective and entertaining tool for vocabulary learning when compared to gamified learning and conventional approaches. The AI-generated storytelling group had the most significant engagement levels, demonstrating its efficacy in captivating and sustaining student attention. Participants in the AI-generated storytelling group reported increased motivation and enhanced learning in qualitative interviews. The engaging essence of AI-generated storytelling, which provides immersive experiences to language learners, enhances these favorable results (Liu, 2025). Thus, AI-generated storytelling may be considered a successful instrument for enhancing language acquisition and student involvement, presenting a viable alternative to conventional and gamified learning techniques.

The results of the current research have considerable ramifications. The evident advantages of AI-generated storytelling provide a compelling rationale for learners to adopt these tools as a significant enhancement to conventional educational approaches. Platforms like Duolingo, however helpful in enhancing engagement relative to traditional training as our findings suggest, appear less effective than personalized, narrative-driven AI methodologies. This highlights the potential of AI not just as a substitute for repetitive exercises, but as a means of embedding vocabulary within coherent, emotionally resonant, and memorable experiences. As a result, learners may actively seek the incorporation of AI-driven educational tools, recognizing their capacity to make language learning both more enjoyable and more effective.

Nonetheless, actualizing this potential in real classrooms depends on practical conditions that the study does not fully address. The practical applications of this research hinge on educators' ability to implement AI tools easily and affordably, yet the manuscript does not resolve critical barriers such as training costs, lack of reliable infrastructure, or limited access to devices, challenges that are especially acute in public-school contexts in Iran and similar settings. Language educators, while often enthusiastic about innovation, may struggle to integrate complex digital tools without adequate support. Recent research by Pathiranage and Karunaratne (2023) emphasizes that effective technology integration requires not only access to tools but also targeted professional development that helps teachers adapt their pedagogy and evaluate digital learning outcomes critically. To strengthen the study's practical relevance, future iterations should include specific, scalable teacher training protocols. For instance, a blended model could combine a short introductory workshop, such as 3 h focused on platform navigation and pedagogical alignment, with ongoing peer mentoring during implementation. Moreover, a basic cost assessment would greatly enhance feasibility planning. While the AI platform was provided free for research purposes, real-world adoption would likely involve expenses such as per-student licensing fees, estimated at 5 to 10 US dollars annually, as well as data usage and technical maintenance. Including such details would help schools and policymakers assess viability and allocate resources accordingly. Policy measures should therefore go beyond curriculum integration and ethical guidelines to also support infrastructural readiness and teacher capacity building. Only through this holistic approach can AI-generated storytelling become a genuinely accessible and equitable resource for EFL learners, rather than a privilege limited to well-resourced institutions.

7. Limitations and directions for future research

When interpreting the study's findings, several limitations should be acknowledged. First, the participants were all Iranian, aged between 20 and 27, and at a B1 proficiency level. This narrow demographic profile limits the extent to which the results can be generalized to learners of different ages, proficiency levels, or cultural backgrounds. Future studies would benefit from including more diverse groups of learners to determine whether the observed effects hold across varied linguistic and educational contexts. Second, the study used only Duolingo as the gamified learning platform. Because Duolingo incorporates narrative elements into its design, it is difficult to separate the impact of gamification from the influence of storytelling when comparing it with the AI-generated narratives. This overlap may have affected how participants responded to each condition. To obtain clearer insights, future research could compare multiple gamified platforms or develop custom gamified tasks that control for narrative features, allowing authors to isolate the specific contribution of gamification to vocabulary learning. Third, the intervention lasted only four weeks. Although this timeframe was sufficient to detect immediate learning gains, it does not allow for a thorough assessment of long-term retention. The study included an eight-week delayed post-test, but longer follow-up periods extending over several months would provide a more accurate picture of whether the benefits of AI-generated storytelling endure. Longitudinal designs are therefore needed to evaluate the sustainability of these instructional approaches.

Fourth, the qualitative data were drawn from interviews with just 15 participants. While the responses revealed consistent themes, this sample size may not have been large enough to reach full data saturation. Future qualitative investigations should aim for a more comprehensive representation of learner perspectives, possibly through iterative sampling and continued data collection until saturation is achieved, in accordance with established practices in applied linguistics research. Last but not least, the criticisms raised about gamified learning in this study appear to stem more from the specific design of Duolingo than from gamification as a pedagogical concept. Moreover, the study does not address important ethical issues related to the use of AI in language learning, such as algorithmic bias or the need for methods to verify the authenticity and reliability of AI-generated content. These concerns are increasingly relevant in digital language education. Future research should therefore incorporate ethical frameworks that examine fairness, transparency, and accountability in AI-driven tools, and explore models that combine human judgment with AI capabilities to support both effective and responsible language learning.

CRediT authorship contribution statement

Ehsan Namaziandost: Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis. **Fidel Çakmak:** Writing – review & editing, Validation, Methodology, Investigation.

Ethics and open data statement

This study did not involve human or animal experimentation requiring formal ethical approval under institutional or national regulations. As a *minimal-risk educational study*, no IRB approval was sought or issued. Informed consent was obtained from all participants, who were fully briefed on the study's objectives, their right to withdraw without penalty, and the safeguards for confidentiality and anonymity. All data were anonymized and securely stored, ensuring adherence to recognized ethical principles of respect for persons, beneficence, and justice. The datasets generated and analyzed during the current study are not publicly available due to privacy and confidentiality agreements with participants. However, they are available from the corresponding author on reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the author(s) made use of ChatGPT, a large language model developed by OpenAI and Grammarly, to assist with refining grammar, enhancing readability, and improving textual clarity. Following the use of this tool, the author(s) thoroughly reviewed, edited, and verified all content to ensure accuracy, originality, and alignment with academic standards. The author(s) take full responsibility for the integrity and final version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abuhassna, H., Yahya, N., Zakaria, M. A. Z. M., Al-Maatouk, Q., & Awae, F. (2021). Guidelines for designing distance learning courses via Moodle to enhance students satisfaction and achievements. *International Journal of Information and Education Technology, 11*(12), 574–582. https://doi.org/10.18178/ijiet.2021.11.12.567

Alhamami, M. (2016). Vocabulary learning through audios, images, and videos: Linking technologies with memory. *Computer-Assisted Language Learning Electronic Journal, 17*(2), 87–112. https://callej.org/index.php/journal/article/view/229.

Anane, C. (2024). Impact of a game-based tool on student engagement in a foreign language course: A three-term analysis. *Frontiers in Education, 9*, Article 1430729. https://doi.org/10.3389/feduc.2024.1430729

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology, 44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Bandura, A. (2001). Social cognitive theory and clinical psychology. In N. J. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 14250–14254). Elsevier. https://doi.org/10.1016/b0-08-043076-7/01340-1.

Beguš, N. (2024). Experimental narratives: A comparison of human crowdsourced storytelling and AI storytelling. *Humanities and Social Sciences Communications, 11*(1). https://doi.org/10.1057/s41599-024-03868-8. Article 1.

Belda-Medina, J., & Goddard, M. B. (2024). AI-driven digital storytelling: A strategy for creating English as a Foreign Language (EFL) materials. *International Journal of Linguistics Studies, 4*(1), 40–49. https://doi.org/10.32996/ijls.2024.4.1.4

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Bhutoria, A. (2022). Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model. *Computers and Education: Artificial Intelligence, 3*, Article 100068. https://doi.org/10.1016/j.caeai.2022.100068

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, 18*(1), 1–21. https://www.jstor.org/stable/1343711.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. *Journal of the Association for Computing Machinery, 37*(4). https://doi.org/10.48550/arXiv.2303.04226. Article 111.

Castro, E., & Shelton-Strong, S. J. (2024). Exploring emotions in language learning: Learners' self-awareness, personal growth, and transformation on a CLIL course. *Language Teaching Research*. https://doi.org/10.1177/13621688241267366. Advance online publication.

Chen, G. (2024). The impacts of AI-driven storytelling applications on language acquisition and literacy development in early childhood education: A systematic review. *Asian Journal of Advanced Research and Reports, 18*(8), 1–12. https://doi.org/10.9734/ajarr/2024/v18i11770

Cho, Y. R., & Yim, D. (2024). The effect of online bilingual shared book reading on vocabulary learning for Korean-English bilingual children. *Communication Sciences & Disorders, 29*(1), 14–27. https://doi.org/10.12963/csd.23999

Chowdhury, M., Dixon, L., Kuo, L., Donaldson, J. P., Eslami, Z., Viruru, R., & Luo, W. (2024). Digital game-based language learning for vocabulary development. *Computers and Education Open, 6*, Article 100160. https://doi.org/10.1016/j.caeo.2024.100160

Chubb, J., Missaoui, S., Concannon, S., Maloney, L., & Walker, J. A. (2022). Interactive storytelling for children: A case-study of design and development considerations for ethical conversational AI. *International Journal of Child-Computer Interaction, 32*, Article 100403.

Colella, S. (2025). "The language of the digital air": AI-generated literature and the performance of authorship. *Humanities, 14*(8). https://doi.org/10.3390/h14080164. Article 164.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.

Csikszentmihalyi, M., & Larson, R. (2014). *Flow and the foundations of positive psychology*. Springer.

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education, 14*(1). https://doi.org/10.1186/s41239-017-0042-5. Article 9.

Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation*. Routledge.

Elaish, M. M., Shuib, L., Hwang, G. J., Ghani, N. A., Yadegaridehkordi, E., & Zainuddin, S. Z. (2023). Mobile English language learning: A systematic review of group size, duration, and assessment methods. *Computer Assisted Language Learning, 36*(3), 430–456. https://doi.org/10.1080/09588221.2021.1931341

Fan, L., & Yao, G. (2025). Cognitive load scale for AI-assisted L2 writing: Scale development and validation. *Frontiers in Psychology, 16*, Article 1666974. https://doi.org/10.3389/fpsyg.2025.1666974

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Fotaris, P., Mastoras, T., Leinfellner, R., & Rosunally, Y. (2016). Climbing up the leaderboard: An empirical study of applying gamification techniques to a computer programming class. *Electronic Journal of E-Learning, 14*(2), 94–110. https://files.eric.ed.gov/fulltext/EJ1101229.pdf.

Fu, J. S., Yang, S. H., & Yeh, H. C. (2022). Exploring the impacts of digital storytelling on English as a foreign language learners' speaking competence. *Journal of Research on Technology in Education, 54*(5), 679–694. https://doi.org/10.1080/15391523.2021.1911008

Godwin-Jones, R. (2019). In a world of SMART technology, why learn another language? *Journal of Educational Technology & Society, 22*(2), 4–13. https://www.jstor.org/stable/26819613.

Grabner-Hagen, M. M., & Kingsley, T. (2023). From badges to boss challenges: Gamification through need-supporting scaffolded design to instruct and motivate elementary learners. *Computers and Education Open, 4*, Article 100131. https://doi.org/10.1016/j.caeo.2023.100131

Han, A., & Cai, Z. (2023). Design implications of generative AI systems for visual storytelling for young learners. In *Proceedings of the 22nd annual ACM interaction design and children conference* (pp. 470–474). Association for Computing Machinery.

Har, F. (2022). English language learning in response to the COVID-19 pandemic: Hong Kong English as second language students' perceptions of Badaboom. *Frontiers in Education, 7*, Article 966059. https://doi.org/10.3389/feduc.2022.966059

Har, F. (2023). Teaching English as a second language in the midst of a paradigm shift: An exploration of students' and teachers' perception of ChatGPT. In *Critical reflections on ICT and education: Selected papers from the HKAECT 2023 International Conference* (pp. 21–35). Singapore: Springer Nature.

Hava, K. (2021). Exploring the role of digital storytelling in student motivation and satisfaction in EFL education. *Computer Assisted Language Learning, 34*(7), 958–978. https://doi.org/10.1080/09588221.2019.1650071

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research & Development, 68*(4), 1875–1901. https://doi.org/10.1007/s11423-020-09807-z

Jia, Y., Xu, B., Karanam, Y., & Voida, S. (2016). Personality-targeted gamification: A survey study on personality traits and motivational affordances. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2001–2013). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858515.

Keller, J. M. (1979). Motivation and instructional design: A theoretical perspective. *Journal of Instructional Development, 2*(4), 26–34. https://doi.org/10.1007/BF02904345

Kingsley, T. L., & Grabner-Hagen, M. M. (2018). Vocabulary by gamification. *The Reading Teacher, 71*(5), 545–555. https://doi.org/10.1002/trtr.1645

Kohnke, L., & Moorhouse, B. L. (2022). Using Kahoot! to gamify learning in the language classroom. *RELC Journal, 53*(3), 769–775. https://doi.org/10.1177/00336882211040270

Koraishi, O. (2024). The intersection of AI and language assessment: A study on the reliability of ChatGPT in grading IELTS writing task 2. *Language Teaching Research Quarterly, 43*, 22–42.

Kristiawan, D., Ferdiansyah, S., & Picard, M. (2022). Promoting vocabulary building, learning motivation, and cultural identity representation through digital storytelling for young Indonesian learners of English as a foreign language. *Iranian Journal of Language Teaching Research, 10*(1), 19–36. https://files.eric.ed.gov/fulltext/EJ1329985.pdf.

Kuo-Ping, L., Shu-Ju, D. T., & Chen-Chung, L. (2018). Enhancing language learning through creation: The effect of digital storytelling on student learning motivation and performance in a school English course. *Educational Technology Research & Development, 66*(4), 913–935. https://doi.org/10.1007/s11423-018-9592-z

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior, 71*, 508–515. https://doi.org/10.1016/j.chb.2015.08.008

Li, D., & Jianxing, W. (2025). The effect of gamified learning monitoring systems on students' learning behavior and achievement: An empirical study. *Entertainment Computing, 52*, Article 100907. https://doi.org/10.1016/j.entcom.2024.100907

Li, M., Ma, S., & Shi, Y. (2023). Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: A meta-analysis. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1253549. Article 1253549.

Liu, L. (2025). Impact of AI gamification on EFL learning outcomes and nonlinear dynamic motivation: Comparing adaptive learning paths, conversational agents, and storytelling. *Education and information technologies*. Advance online publication. https://doi.org/10.1007/s10639-024-13296-5

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL, 31*(3), 293–311. https://doi.org/10.1017/S0958344019000065

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Nadeem, M., Oroszlanyova, M., & Farag, W. (2023). Effect of digital game-based learning on student engagement and motivation. *Computers, 12*(9), 177. https://doi.org/10.3390/computers12090177

Namaziandost, E., & Rezai, A. (2024). Interplay of academic emotion regulation, academic mindfulness, L2 learning experience, academic motivation, and learner autonomy in intelligent computer-assisted language learning: A study of EFL learners. *System, 125*, Article 103419. https://doi.org/10.1016/j.system.2024.103419

Nami, F., & Asadnia, F. (2024). Exploring the effect of EFL students' self-made digital stories on their vocabulary learning. *System, 120*, Article 103205. https://doi.org/10.1016/j.system.2023.103205

Nguyen-Viet, B., & Nguyen-Viet, B. (2025a). Gamification in Vietnamese education: Assessing psychological need satisfaction, intrinsic motivation, and learning effectiveness. *Learning and Motivation, 89*, Article 102101. https://doi.org/10.1016/j.lmot.2025.102101

Nguyen-Viet, B., & Nguyen-Viet, B. (2025b). The synergy of immersion and basic psychological needs satisfaction: Exploring gamification's impact on student engagement and learning outcomes. *Acta Psychologica, 252*, Article 104660. https://doi.org/10.1016/j.actpsy.2024.104660

Pathiranage, A., & Karunaratne, T. (2023). Teachers' agency in technology for education in pre- and post-COVID-19 periods: A systematic literature review. *Education Sciences, 13*(9). https://doi.org/10.3390/educsci13090917. Article 917.

Piaget, J. (1970). *Science of education and the psychology of the child*. Viking Press.

Ramezanzadeh, A., & Ebadi, S. (2025). AI-generated storytelling for socially just EFL pedagogy: Exploring learner emotions through Fraser's justice theory. *System, 129*, Article 103588. https://doi.org/10.1016/j.system.2024.103588

Rasti-Behbahani, A., & Shahbazi, M. (2022). Investigating the effectiveness of a digital game-based task on the acquisition of word knowledge. *Computer Assisted Language Learning, 35*(8), 1920–1945. https://doi.org/10.1080/09588221.2020.1846567

Reinhardt, J. (2017). Digital gaming in L2 teaching and learning. In C. A. Chapelle, & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 202–216). Wiley-Blackwell.

Rezai, A., Soyoof, A., & Reynolds, B. L. (2024). Disclosing the correlation between using ChatGPT and well-being in EFL learners: Considering the mediating role of emotion regulation. *European Journal of Education, 59*(4). https://doi.org/10.1111/ejed.12752. Article e12752.

Riazi, A. M. (2016). *The routledge encyclopedia of research methods in applied linguistics: Quantitative and statistical analyses*. Routledge.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Ryan, R. M., & Deci, E. L. (2024). Self-determination theory. In *Encyclopedia of quality of life and well-being research* (pp. 6229–6235). Springer International Publishing.

Sadeghi, K., Sağlık, E., Mede, E., Samur, Y., & Comert, Z. (2022). The effects of implementing gamified instruction on vocabulary gain and motivation among language learners. *Heliyon, 8*(11), Article e11811. https://doi.org/10.1016/j.heliyon.2022.e11811

Saleem, A. N., Noori, N. M., & Ozdamli, F. (2022). Gamification applications in e-learning: A literature review. *Technology, Knowledge and Learning, 27*(1), 139–159. https://doi.org/10.1007/s10758-020-09487-x

Samsul, S. A., Yahaya, N., & Abuhassna, H. (2023). Education big data and learning analytics: A bibliometric analysis. *Humanities and Social Sciences Communications, 10*(1). https://doi.org/10.1057/s41599-023-02176-x. Article 1.

Sanosi, A. B. (2018). The effect of Quizlet on vocabulary acquisition. *Asian Journal of Education and e-Learning, 6*(4), 151–158. https://doi.org/10.24203/ajeel.v6i4.5446

Schmoelz, A. (2018). Enabling co-creativity through digital storytelling in education. *Thinking Skills and Creativity, 28*, 1–13. https://doi.org/10.1016/j.tsc.2018.02.002

Setiawan, M. R., & Wiedarti, P. (2020). The effectiveness of Quizlet application towards students' motivation in learning vocabulary. *Studies in English Language and Education, 7*(1), 83–95. https://doi.org/10.24815/siele.v7i1.1

Skinner, B. F. (1965). *Science and human behavior (No. 92904)*. Simon and Schuster.

Smiderle, R., Rigo, S. J., Marques, L. B., de Miranda Coelho, J. A. P., & Jaques, P. A. (2020). The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learning Environments, 7*(1). https://doi.org/10.1186/s40561-019-0098-x. Article 1.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*. Sage Publications.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

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

Waluyo, B., & Balazon, F. G. (2024). Exploring the impact of gamified learning on positive psychology in CALL environments: A mixed-methods study with Thai university students. *Acta Psychologica, 251*, Article 104638. https://doi.org/10.1016/j.actpsy.2024.104638

Waluyo, B., & Bucol, J. L. (2021). The impact of gamified vocabulary learning using Quizlet on low-proficiency students. *CALL-EJ, 22*(1), 158–179. http://callej.org/journal/22_1/waluyo_bucol_2021.pdf.

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System, 127*, Article 103533. https://doi.org/10.1016/j.system.2024.103533

Wu, J., & Chen, D.-T. V. (2020). A systematic review of educational digital storytelling. *Computers & Education, 147*, Article 103786. https://doi.org/10.1016/j.compedu.2019.103786

Yang, W., & Dai, W. (2011). Rote memorization of vocabulary and vocabulary development. *English Language Teaching, 4*(4), 61–68. https://doi.org/10.5539/elt.v4n4p61

Yi, Q., He, Y., Wang, J., Song, X., Qian, S., Yuan, X., … Ni, J. (2025). Score: Story coherence and retrieval enhancement for ai narratives. *arXiv preprint arXiv:2503.23512*.

Zainuddin, Z., Chu, S. K. W., & Perera, C. J. (2024). Gamification in the flipped classroom. In R. N. Landers, & K. N. Bauer (Eds.), *Gamification strategies for retention, motivation, and engagement in higher education: Emerging research and opportunities* (pp. 70–89). Springer. https://doi.org/10.1007/978-981-97-2219-8_4.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zhang, R., Zou, D., & Cheng, G. (2023). Learner engagement in digital game-based vocabulary learning and its effects on EFL vocabulary development. *System, 119*, Article 103173. https://doi.org/10.1016/j.system.2023.103173

Zhou, S. (2024). Gamifying language education: The impact of digital game-based learning on Chinese EFL learners. *Humanities and Social Sciences Communications, 11*(1). https://doi.org/10.1057/s41599-024-04073-3. Article 1.