Digital Storytelling with AI Text-to-Image: A Creative Path to Vocabulary Retention

Le Thi Kieu Van[1*], Van Huynh Ha Le[2]

[1] Van Hien University, Vietnam
[2] Van Lang Language Institute, Foreign Languages Faculty, Van Lang University, Vietnam
[*]Corresponding author's email: vanltk@vhu.edu.vn
[*] https://orcid.org/0009-0008-7856-9359

Received: 31/01/2025 Revision: 01/08/2025 Accepted: 11/08/2025 Online: 20/11/2025

ABSTRACT

This mixed-method study investigates the efficacy of combining AI text-to-image with a digital storytelling approach to facilitate vocabulary retention. The motivation behind this research stems from the challenges many students face in retaining new words and effectively employing them in real-life communication. The study also reveals students' perceptions of AI text-to-image tools aiding word retention. The research was conducted over 6 weeks and involved 80 second-year students divided into two groups: a control group and a treatment group. Data collection instruments included pre- and post-tests assessing targeted vocabulary, comic books - storytelling rubrics, and open-ended questionnaires. The findings show that students who experienced AI text-to-image tools exhibited positive reactions and emotions towards digital storytelling AI text-to-image activities. The experimental group improved more in both immediate and delayed vocabulary retention than the control group. Despite the short intervention time, AI text-to-image combined with DST offers a new, inspiring strategy for vocabulary learning. The study recommends pedagogical and future actions for English language teaching and learning.

Keywords: digital storytelling, AI text-to-image, vocabulary acquisition, vocabulary retention

Introduction

Vocabulary acquisition is crucial in second language acquisition because a rich vocabulary repertoire facilitates communication and success in language learning (Nation, 2001; Schmitt, 2008). A larger vocabulary enables better comprehension and more fluent, accurate communication. However, Webb (2005) believed that vocabulary retention remains one of the barriers language learners face when acquiring a new language, regardless of its significance. The reasons for these difficulties were learners' struggle with rote memorization of context-separated words (Nation, 2001; Yang & Dai, 2011), limited exposure, like the Vietnam EFL case, and lack of practice (Schmitt, 2010). Furthermore, cognitive overload during language learning, where learners are required to process grammar, pronunciation, and meaning simultaneously, can hinder long-term retention of vocabulary (Baddeley, 1997).

In Vietnam, these challenges are particularly pronounced due to traditional teaching

CITATION | Le, T. K. V., & Le, V. H. H. (2025). Digital Storytelling with AI Text-to-Image: A Creative Path to Vocabulary Retention. *Computer-Assisted Language Learning Electronic Journal (CALL-EJ)*, 26(4), 443-462.

methodologies emphasizing passive memorization over active engagement (Tran, 2013; Nguyen & Jaspaert, 2021; Long & Van, 2019). Vocabulary learning activities in the Vietnamese context have followed a teacher-led approach, and the shortage of interactivity for the practical use of newly learned words led to a low retention rate (Vu & Peters, 2021). In this problematic context, integrating AI-generated images with digital storytelling (DST) offers a novel solution to enhance vocabulary retention among English learners. AI text-to-image tools like DALL-E generate dynamic, personalized visuals, unlike traditional DST, which uses static images. This study brings together AI text-to-image tools and digital storytelling (DST) to create a multimodal and adaptive learning environment, addressing gaps in earlier research where this link has not been fully explored. The approach supports both cognitive and affective learning by offering comprehensible input in line with Krashen's i+1 hypothesis. Grounded in constructivist and multimodal theories, DST provides a creative space for learners to connect vocabulary with images, text, and group narratives (Robin, 2008). Integrating AI-generated visuals with DST helps learners picture word meanings and strengthen retention.

The study focuses on two aims:
(i) to examine how combining DST with AI text-to-image tools affects immediate and delayed vocabulary retention, and
(ii) to investigate students' views on using these tools in their vocabulary learning.

Literature review

Vocabulary Retention: Definitions and Previous Research

Vocabulary retention is the capacity of language learners to recall and use words they have already learned over time (Nation, 2013). Most scholars view it as a process that unfolds in two main stages. Immediate retention refers to recalling vocabulary shortly after learning, typically measured right after instruction or intervention. Delayed retention evaluates how well learners retain vocabulary after a more extended period (e.g., days, weeks, or months), focusing on the durability of learning (Al-Obaydi & Pikhart, 2024; Agnes & Srinivasan, 2024).

Definitions of vocabulary retention may vary in terms of emphasis. Some focus on cognitive mechanisms, others on instructional strategies, and others on emotional and motivational dimensions. For instance, Ahmad (2019) emphasizes retention as being influenced by multimodal learning inputs. Hill (2022) and Zabidin (2015) view retention as linked to emotional and motivational factors, such as humor and mnemonic devices, suggesting that retention is cognitive and affected by affective engagement. Allanazarova (2020) frames retention within meaningful learning theory, where anchoring vocabulary to existing cognitive structures promotes better long-term recall.

Vocabulary Acquisition versus Vocabulary Retention

Nation (2013) emphasizes vocabulary acquisition as a comprehensive process involving learning, understanding, and using words. Unlike vocabulary acquisition (initial learning), retention is a sub-category of acquisition, which is a cumulative psychological review and belongs to long-term storage and recall. Baisel and Ramachandran (2024) point out the two processes as a continuum. Without the retention process, attempts to acquire words become useless since learners cannot apply what they learn to practical usage. Hence, successful communication as well as language proficiency are based on vocabulary retention success (Schmitt, 2020; Webb, 2005).

Previously Employed Strategies to Improve Vocabulary Retention

Diverse techniques have been used in English teaching to support vocabulary retention. The reviewed studies collectively illustrate a clear evolution in vocabulary retention research, transitioning from traditional rote memorization toward dynamic, personalized, and technology-enhanced strategies.

Ahmad (2019) investigated multimedia glosses as tools for vocabulary acquisition and retention among EFL learners. The experimental group interacted with the study's glossed texts featuring text, images, and videos, while the control group read unglossed texts. Results showed significantly higher scores in the experimental group's immediate and delayed vocabulary retention tests, emphasizing the dual benefit of enhanced short-term learning and sustained long-term recall. The findings align with the Dual Coding Theory, which posits that multimodal input, visual and verbal, facilitates deeper cognitive processing. Multimedia glosses reach a wider range of learners than traditional ones, as they appeal to both visual and linguistic strengths.

In one study, Alsadoon (2021) tested an AI chatbot equipped with a dictionary, a concordancer, and L1 translation. The dictionary proved most effective for short-term recall because learners could look up meanings instantly, while translation into the first language gave a small advantage in long-term retention. This reflects the interactionist view that learning is shaped by immediate engagement. The method is more interactive than static glosses, though it lacks the richness of image-based tools. Nematollahi et al. (2017) showed that visual storytelling supports retention better than oral storytelling. Pictures strengthened short-term recall, while context supported memory over time. Their findings confirm dual coding theory and point to the role of context in vocabulary learning.

Ashkan and Seyyedrezaei (2016) investigated corpus-based teaching. Learners worked with authentic texts, and repeated exposure improved short-term memory. Inferring meaning from real examples aided long-term recall. Like glosses and storytelling, this method stresses context, but it differs by promoting discovery learning, with students analyzing language directly.

Unlike other strategies, Hill (2022) examined various mnemonic strategies, such as acronyms, rhymes, and visual mnemonics, for vocabulary retention in the Chinese EFL context. Mnemonics concentrate on cognitive strategies: metacognitive skills and learner autonomy. Thanks to their engaging and structured nature, mnemonics improve immediate retention. Meanwhile, their deeper semantic connections (keyword method) support delayed retention. Compared to corpus-based and AI tools, mnemonic strategies encourage learners to have more autonomy.

Zabidin (2015) suggested using humor as a teaching tool to help Malaysian EFL learners remember new vocabulary more effectively. The study found that students retained words better in the short and long term when exposed to humorous content rather than plain texts. Humor does more than just aid memory; it eases anxiety, keeps learners motivated, and makes the learning experience more enjoyable.

Unlike other authors focusing on remembering and recalling vocabulary, Allanazarova (2020) applied cognitive theories to address forgetting in vocabulary learning, emphasizing the importance of meaningful learning over rote memorization. The findings reveal that combined pictures, texts, and stories will boost immediate and delayed retention. This reaches a consensus with Ausubel’s meaningful learning theory that stresses the attempt to attach new words to existing cognitive frameworks.

Al-Obaydi and Pikhart (2024) revisited Total Physical Response (TPR), shedding light on its role in connecting movement with vocabulary learning. Their study focused on secondary school students and introduced the "spell the word in action" game, where learners physically enact words to reinforce retention. The findings showed that TPR has a lasting effect. Words learned in class were still remembered weeks later. TPR is not only a classroom activity but also a way to make teachers, students, and researchers think again about usual methods and consider new ones for teaching vocabulary.

Baisel and Ramachandran (2024) tested AI-made mnemonic keywords with Anki flashcards. These keywords helped intermediate learners recall vocabulary both soon after learning and later on. Because the keywords were matched to different learning styles, they gave students more support and flexibility. This shows that artificial intelligence can add value to traditional study practices. Nation (2013) is well known for work on vocabulary. He focused on cumulative learning, meaning-focused input, and psychological factors in retention. He divided words into high-, mid-, and low-frequency groups, pointed out the role of collocations, and explained how to balance word exposure with context. Although he did not test immediate versus delayed recall directly, his stress on spaced repetition and review supports long-term learning.

Tran (2013) described cultural and system issues in Vietnam's higher education that encourage passive learning, especially rote memorization. His study points to the need for more learner-centered ways to build vocabulary. Moelyono et al. (2023) looked at how learners used Google Translate. It gave quick access to meanings, reduced anxiety, and built confidence. Still, too much dependence reduced deeper use of the target language. This suggests that while technology is useful, real vocabulary learning needs both digital help and active practice.

Jamal et al. (2024) focused on rote repetition, but newer research highlights multimodal strategies that use more than one sense (Ahmad, 2019; Nematollahi et al., 2017). Videos and stories, for example, support dual coding theory by linking words with both images and text. Other active methods include TPR (Al-Obaydi & Pikhart, 2024), AI for personalized learning (Alsadoon, 2021), corpus-based study (Ashkan & Seyyedrezaei, 2016), and affective approaches such as humor (Zabidin, 2015) or AI mnemonics (Baisel & Ramachandran, 2024). All these move learning away from rote memory toward creativity and emotional involvement. Yet few studies compare these methods or combine them. The use of AI-generated images with digital storytelling is still new, and this study explores that gap.

Digital Storytelling and Its Rationales

Digital storytelling (DST) brings together text, sound, images, and video to make stories for learning. Research shows it can improve language skills, increase motivation, and support critical thinking by connecting traditional storytelling with digital tools (Lim et al., 2022; Albishi & Alqiawi, 2022; Barua, 2023).

Lim et al. (2022) describe DST as building stories through digital media such as text, pictures, and audio. Albishi and Alqiawi (2022) focus on vocabulary learning and show how images, gestures, and context help learners understand meaning. Barua (2023) notes that DST combines technology and creativity to support vocabulary growth, critical thinking, and cultural learning. Sembiring and Simajuntak (2023) view DST as a modern alternative to traditional methods, especially useful for vocabulary teaching in EFL classrooms.

Although each author looks at a different side, vocabulary, thinking skills, culture, or inclusion, they all agree that DST uses many modes and keeps learners engaged. Albishi and Alqiawi (2022) show that multimedia helps students remember words. Barua (2023) highlights DST's role in developing critical thinking and intercultural awareness. Le (2020)

finds that it also improves enjoyment and oral fluency. Belda-Medina and Goddard (2024) explore how DST supports marginalized groups, while Poitras-Pratt (2020) shows its use in preserving Indigenous culture.

Over time, DST has grown from a classroom method into a broader, flexible approach with both educational and social value. Its strength is in combining storytelling with visuals and interaction to support expression, inclusion, and lasting learning.

AI Text-to-Image: A Newly-Emergent Technology

Text-to-image AI systems rely on machine learning models such as GANs and diffusion models to turn written input into pictures. This study focuses on DALL-E because it produces high-quality and accurate images, as shown by its low Fréchet Inception Distance (FID) of 9.0% and Peak Signal-to-Noise Ratio (PSNR) of 9.88 (Jamal et al., 2024). These results show that DALL-E can generate realistic and context-appropriate visuals, which supports the study's goal of improving vocabulary learning.

When compared with other systems like Google Imagen, GROK, or Stable Diffusion, DALL-E fits classroom use better by providing more precise pictures that aid memory. In language learning, it helps illustrate abstract ideas and strengthen word–image links, which supports multimodal learning (Żelaszczyk & Mańdziuk, 2024). For example, turning a phrase such as *mysterious legend* into a picture can improve both understanding and recall.

Although often applied in digital art, text-to-image AI also shows strong promise in education by giving learners personalized visuals that carry meaning. This matches Krashen's Input Hypothesis (1982), since the images can provide comprehensible input at the i+1 level. Yet, some problems remain. These include limits in handling complex prompts, issues of bias and copyright, and dependence on user skill (Jamal et al., 2024; Żelaszczyk & Mańdziuk, 2024). Lower-cost tools such as Stable Diffusion offer alternatives, though sometimes at the expense of quality. Current measures like FID and PSNR also do not fully capture human judgment, pointing to the need for better evaluation methods in education.

Research Gaps

Although vocabulary retention, digital storytelling (DST), and AI text-to-image tools have been widely studied, little work has integrated these elements to enhance vocabulary retention. This study seeks to fill that gap. While DST uses text, audio, images, and video to enrich language learning (Lim et al., 2022; Albishi & Alqiawi, 2022), most implementations rely on static or pre-made visuals, limiting personalization. AI tools like DALL-E, which generate custom images from text, offer a dynamic alternative (Jamal et al., 2024). Yet, the combination of DST with AI-generated visuals for vocabulary learning remains underexplored.

Research often isolates the cognitive or emotional aspects of retention—mnemonics (Hill, 2022) or humor (Zabidin, 2015)—without considering their interaction. DST can integrate affective and contextual elements to support memory (Barua, 2023), and adding AI-generated imagery could further enhance both emotional engagement and cognitive processing. Although Żelaszczyk and Mańdziuk (2024) highlight the role of AI visuals in multimodal learning, their use within narrative-based strategies like DST is still an open area.

Current vocabulary instruction also underutilizes adaptive, personalized learning. DST tends to employ static media that doesn't address individual needs, whereas AI tools can generate visuals tailored to each learner. Their integration could support differentiated instruction, a potential largely untapped in existing research.

Finally, few studies address the relationship between immediate and long-term retention (Al-Obaydi & Pikhart, 2024; Ahmad, 2019). DST can frame initial learning, while AI visuals provide contextual reinforcement to support long-term recall. Together, these tools may bridge the gap between short-term acquisition and sustained vocabulary retention.

Theoretical Foundations Supporting Digital Storytelling in Vocabulary Retention

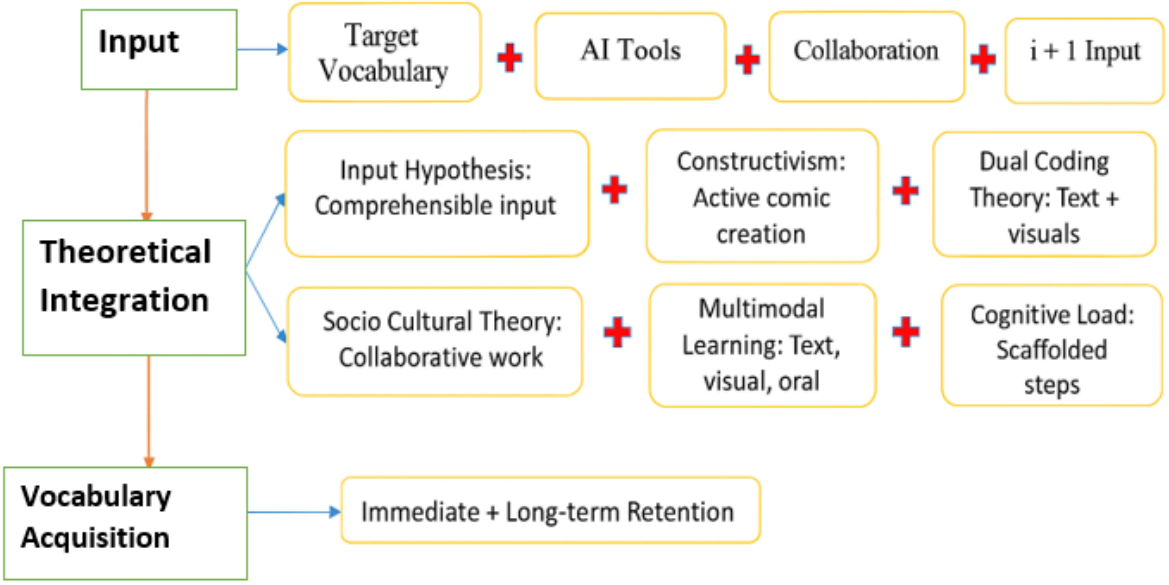
Several learning theories support digital storytelling (DST) as an effective tool for vocabulary retention. Constructivism emphasizes learners' active role in building knowledge through meaningful, collaborative engagement. In DST, making digital comic books allows students to use new vocabulary in real stories. This hands-on and group-based work supports memory through social interaction (Piaget, 1972; Vygotsky, 1978).

Dual Coding Theory explains that learning improves when text and images are combined, since both verbal and visual channels are used (Paivio, 1971, 1990). DST applies this idea by linking words to pictures, which helps recall and understanding. Multimodal Learning also stresses the use of different modes. By drawing, writing, and speaking, students experience multiple ways of learning that fit different styles and improve memory (Mayer, 2002). Sociocultural Theory sees learning as a social process. When learners co-create digital stories, they share ideas, guide one another, and practice vocabulary through interaction and joint meaning-making (Vygotsky, 1978). Showing their stories to others adds context, which strengthens communicative ability. Cognitive Load Theory reminds us that learning requires careful management of mental effort. DST breaks the work—designing, creating, presenting—into smaller steps. This reduces extra load and allows students to focus more on vocabulary (Sweller, 1988).

Taken together, these theories present DST as an engaging, structured, and collaborative way to learn vocabulary. This study builds on that base by combining DST with AI text-to-image tools to give learners a multimodal and personalized approach suited to today's AI-supported classrooms. Figure 1 shows the framework for this study.

Figure 1

Conceptual Framework: AI-Enhanced Digital Storytelling Framework for Vocabulary Acquisition



Krashen's Input Hypothesis (1982) holds that language acquisition occurs best when learners receive comprehensible input just beyond their current level (i+1). Digital storytelling facilitates this by presenting vocabulary through meaningful, real-life narratives. AI text-to-image tools like DALL-E enhance the process by generating visuals aligned with learners' proficiency, supporting the i+1 principle. For instance, illustrating the word "heroic" in a story about "Heroes" makes abstract vocabulary more accessible and contextually grounded. Collaborative storytelling also fosters negotiation of meaning and peer scaffolding, further reinforcing comprehensible input.

By integrating the Input Hypothesis with multimodal learning and constructivist principles, this study aims to support vocabulary retention in an engaging, learner-centered environment. AI streamlines visual creation, addressing earlier challenges in manual comic design (Le & Doan, 2023). The result is faster, more accurate, and personalized visuals that support memory. However, over-reliance on AI may reduce learner creativity. To mitigate this, teachers should provide tutorials on AI tools and prompt students to evaluate and improve AI-generated images, ensuring sustained engagement and active learning.

Research Questions

The study addresses the following research questions:

1. Does integrating digital storytelling and AI text-to-image tools improve immediate and delayed vocabulary retention among second-year English language learners?

2. What are students' perceptions of AI text-to-image tools supporting vocabulary retention?

This research offers insights into a novel method for addressing vocabulary learning challenges in Vietnam, contributing to both theory and practice in language education.

Methods

Pedagogical Setting & Participants

The study used non-probability convenience sampling to recruit participants. Two second-year English major classes were selected based on their availability, resulting in 80 students, 40 in the control group and 40 in the experimental group. Participants, aged 18 to 20, were enrolled in a Reading course and had intermediate English proficiency. All students completed pre- and post-tests to assess vocabulary retention and responded to open-ended questionnaires about their perceptions of the learning tools and methods. This sampling method enabled efficient recruitment while ensuring all participants contributed to both quantitative and qualitative components of the study (Creswell & Creswell, 2017).

Design of the Study

This study uses a convergent mixed-method design (also called concurrent), where quantitative and qualitative data are gathered and analyzed at the same time. The two strands are then brought together during interpretation to give a fuller view of the research questions (Creswell & Plano, 2017). In this study, test scores on vocabulary retention are combined with students' reflections to provide both numerical and descriptive insights.

The data come from three main sources: pre- and post-tests (for delayed retention), four comic books and four storytelling presentations from each group (for immediate retention), and open-ended questionnaires. To measure delayed retention, pre- and post-tests were given, each with 20 items covering the same 20 target words taught across four weeks. Each test had 10 multiple-

choice questions and 10 gap-filling items, which helped limit guessing and gave a clearer picture of word knowledge.

Immediate retention was checked through student-made comic books and in-class storytelling. Their work was scored on the correct use of target words, clarity of the story, grammar, and presentation skills. For student perceptions of AI image tools, open-ended questionnaires with three guiding questions were used. These students were asked about their experience with the tools, the process of making comic books, and whether they would like to continue using this method.

To secure reliability and validity, two experts reviewed all instruments, following best practices for tool development (Creswell & Plano, 2017). By combining test data with student feedback, the study provides a broad evaluation of the intervention's effectiveness and learners' experiences with it.

Data collection

In the experimental group, data collection began in week 1 with a tutorial session. Students were shown how to use DALL-E to generate images and Canva to put together comic books. Standardized training ensured that images were of consistent quality and matched the lesson content, reducing possible confounding factors. The tutorial included guidance on writing scripts with the target words, producing images, and compiling them into comics.

Also in week 1, students took a pre-test to measure their starting level of vocabulary knowledge. From weeks 2 to 5, students engaged in reading-based lessons on topics such as A Book of Secrets, Mysterious Legends, Heroes, and Friendship. Each session began with a DALL-E-generated model story, followed by explicit vocabulary instruction and comprehension checks. Students then worked in groups of five to create their own stories using DALL-E, and group representatives presented their work at the end of each lesson. Figure 2 displays sample pages from the DALL-E-generated comic used as the teacher's model in Unit 1 (Book of Secrets).

Figure 2

Sample Comic Pages Featuring Targeted Vocabulary from the DALL-E-Generated Model (Unit 1: Book of Secrets)

In contrast, the control group followed a traditional teaching method, where the teacher introduced vocabulary through direct instruction of word meanings and forms, followed by group-based comprehension checks. At the end of each lesson, students worked in eight groups

to answer teacher-led questions, with scores assigned based on group responses. After completing the four lessons, both groups took a post-test to assess vocabulary retention and completed open-ended questionnaires in week 6 to reflect on their learning experiences. Table 1 summarizes the data collection timeline and key activities for both groups.

Table 1

Data Collection Process and Theoretical Alignment in the Study

Stage	Description	Theories Aligned
Tutorial Session (Week 1)	Training in DALL-E and Canva, and learning to create comic books.	- Constructivism (active creation of comics) - Multimodal Learning (visual + text integration) - Cognitive Load Theory (scaffolding AI use).
Pre-Test (Week 1)	Assessment of initial vocabulary knowledge with a 20-question test (10 multiple-choice, 10 gap-fill).	
Lessons on Reading Topics (Weeks 2-5)	Interactive lessons with topics (Book of Secrets, Mysterious Legends, Heroes, Friendship)	- Constructivism - Multimodal Learning - Cognitive Load Theory - Input Hypothesis - Sociocultural Theory - Dual Coding Theory
	- Teacher provides DALL-E-generated model story for group reading. - Teacher explains vocabulary and checks comprehension.	- Input Hypothesis (comprehensible input via DALL-E-generated stories) - Multimodal Learning (text, visuals, oral).
	- Students create their own stories in groups using DALL-E.	- Constructivism (creating AI stories), Dual Coding Theory (text + images) - Sociocultural Theory (group collaboration and discussion of ideas) - Multimodal Learning (text, visuals, oral).
	- Groups present stories to the class at the end of the lesson.	Sociocultural Theory (group collaboration and presentations).
Post-Test (Week 6)	Assessment of vocabulary retention with a 20-question test (same format as pre-test).	
Open-Ended Questionnaire (Week 6)	Students answer three open-ended questions to reflect on their experience with DALL-E and storytelling.	

Data Analysis

The data set included quantitative and qualitative components to address the research questions comprehensively. Quantitative data gathered from the pre-test and post-test scores were analyzed using Independent Samples T-tests in IBM SPSS (version 26) to compare the vocabulary retention levels between the experimental and control groups. Cohen's d was also

calculated to measure the effect size, providing insights into the magnitude of the intervention's impact. For immediate vocabulary retention and storytelling skills, a rubric (as shown in Table 2) was used to evaluate the number of targeted vocabulary words used, the story message, grammar, and presentation skills. The rubric underwent an expert review process to ensure its validity. However, for the specific comparison of immediate vocabulary retention, only the scores for the number of targeted vocabulary words used were analyzed using Independent Samples T-tests.

Table 2

Rubric for Evaluation of Immediate Vocabulary Retention and Storytelling Skill

Criteria	Excellent (9–10)	Good (7–8)	Fair (5–6)	Needs Improvement (1–4)
Targeted Vocabulary (10 pts)	Uses all targeted vocabulary appropriately in context; demonstrates creative and varied usage in the story.	Uses most targeted vocabulary correctly, but some may lack contextual appropriateness or variety.	Limited use of targeted vocabulary; some errors in usage or understanding of meaning.	Minimal or incorrect use of targeted vocabulary; shows little understanding of the words.
Story Message (10 pts)	The story is highly creative, coherent, and engaging; the message is clear and aligns well with the lesson topic.	The story is clear and coherent, with good creativity, though the message may not be completely aligned with the topic.	The story is somewhat clear but may lack coherence, creativity, or alignment with the lesson topic.	The story is unclear and lacks coherence, creativity, or relevance to the lesson topic.
Grammar (10 pts)	Grammar is consistently accurate with minor or no errors; sentences are varied and well-constructed.	Grammar is mostly accurate, with occasional errors that do not hinder meaning; sentence structure is adequate.	Frequent grammar errors that may interfere with meaning; sentence structure is basic or repetitive.	Grammar errors are pervasive and significantly interfere with meaning; sentences are poorly constructed.
Presentation (10 pts)	The presentation was confident, engaging, and well-organized; all group members contributed effectively.	The presentation is organized and straightforward, but confidence or engagement may vary; most group members contribute.	The presentation is somewhat organized but lacks clarity, engagement, or balance in group participation.	The presentation is disorganized, unclear, and lacks engagement; there is limited contribution from group members.

Braun and Clarke's (2006) six-step thematic analysis was applied to the qualitative data from open-ended questionnaires. The steps included: (1) familiarization through repeated reading,

(2) generating initial codes, (3) identifying themes, (4) reviewing theme accuracy, (5) defining and naming themes, and (6) producing a thematic report aligned with the research questions. This method captured students' perspectives on how DALL-E supports vocabulary retention. 'Inconsistencies' were grouped under 'Limitations in AI Outputs' to refine the thematic structure (see Table 5). The study combined this qualitative approach with quantitative methods to offer both statistical precision and interpretive depth.

Results

Both quantitative and qualitative data were analyzed to answer two questions:

1. Does integrating digital storytelling and AI text-to-image tools improve immediate and delayed vocabulary retention among second-year English language learners?

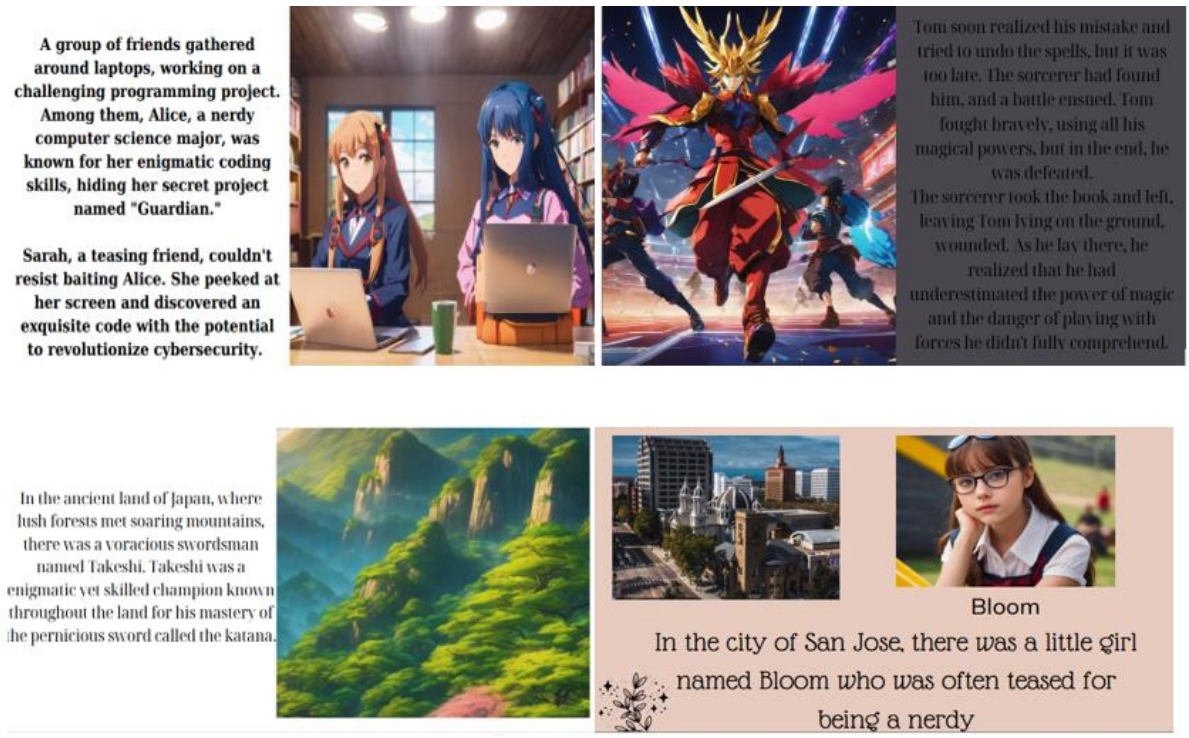
2. What are students' perceptions of AI text-to-image tools supporting vocabulary retention?

Immediate Vocabulary Retention Results

After four weeks, 32 digital comic books were collected from the eight experimental groups to investigate and score the number of targeted vocabulary words used in each (see Figure 3).

Figure 3

Extracts of Participants' DALL-E Text-To-Image Comics



The control group's scores were recorded based on group answers (8 groups) to comprehension check questions conducted after each lesson. The number of targeted vocabulary words retained by each group was tallied and summarized for both groups. These results are presented in Table 3. This table provides a comparative overview of the retention levels demonstrated by the groups in both instructional settings.

Table 3

Immediate Vocabulary Retention Scores of Experimental and Control Groups

Ex G	Week 1	Week 2	Week 3	Week 4	Con G	Week 1	Week 2	Week 3	Week 4
G1	8	9	7	8	**G1**	7	6	8	7
G2	9	8	8	7	**G2**	6	7	6	6
G3	7	9	7	8	**G3**	8	7	7	8
G4	10	10	9	10	**G4**	6	6	6	7
G5	6	7	6	6	**G5**	5	6	5	5
G6	8	9	8	7	**G6**	7	7	8	7
G7	10	9	9	10	**G7**	8	7	9	8
G8	7	6	7	6	**G8**	6	5	7	6

An Independent Samples T-test was conducted to compare immediate vocabulary retention scores between the experimental and control groups. The results revealed a statistically significant difference, with the experimental group ($M = 7.97$, $SD = 1.33$) scoring higher than the control group ($M = 6.69$, $SD = 1.03$), $t(62) = -4.31$, $p < .001$. The mean difference of 1.28 points, along with a large effect size (Cohen's $d = 0.96$), shows that using digital storytelling with DALL-E led to significantly better vocabulary retention than traditional teaching methods (see Table 4).

Table 4

Descriptive Statistics and T-Test Results for Vocabulary Retention

Group	N	Mean	SD	SE Mean	t	df	p	Mean Difference	95% CI (Lower)	95% CI (Upper)	Cohen's d
Control Group	32	6.69	1.03	0.18	-4.31	62	< .001	-1.28	-1.88	-0.69	0.96
Experimental Group	32	7.97	1.33	0.24							

The results show that students in the experimental group, who used digital storytelling with DALL-E, scored much higher in vocabulary retention than those in the control group. Their mean score ($M = 7.97$, $SD = 1.33$) indicates that the intervention was very effective. The large effect size (Cohen's $d = 0.96$) further confirms the strength of this approach. These findings suggest that combining multimodal and interactive methods can strongly improve language learning outcomes.

Delayed Vocabulary Retention Results

An independent samples t-test was run to compare delayed vocabulary retention between the

experimental and control groups (see Table 5). The pre-test results showed no significant difference, suggesting that both groups started with similar vocabulary levels (t(71.14) = -0.32, p = .750). However, the post-test revealed a statistically significant difference, with the experimental group (M = 7.58, SD = 0.93) scoring higher than the control group (M = 7.05, SD = 0.90), t(78) = -2.56, p = .012. The effect size, Cohen's d = 0.57, indicates a moderate effect of the intervention, suggesting that integrating digital storytelling and DALL-E positively influenced long-term vocabulary retention.

Table 5

Descriptive Statistics and T-Test Results for Delayed Vocabulary Retention

Test	Group	N	Mean	SD	SE Mean	t	df	p	Mean Difference	95% CI (Lower)	95% CI (Upper)	Cohen's d
Pre-Test	Control Group	40	4.60	1.60	0.25	-0.32	71.14	.750	-0.10	-0.72	0.52	0.06
	Experimental Group	40	4.70	1.16	0.18							
Post-Test	Control Group	40	7.05	0.90	0.14	-2.56	78	.012	-0.53	-0.93	-0.12	0.57
	Experimental Group	40	7.58	0.93	0.15							

The results show that using digital storytelling and DALL-E significantly improved delayed vocabulary retention in the experimental group compared to the control group. While no significant difference was observed in the pre-test scores, the post-test results suggest that the intervention had a meaningful and positive impact on long-term vocabulary retention.

Students' Perceptions of AI Text-To-Image Tools

Responses to the open-ended questionnaires were thematically analyzed using Braun and Clarke's (2006) method. Recurring topics related to the second research question (students' perceptions of AI text-to-image tools in vocabulary retention) were identified as codes, grouped into larger categories, and organized into overarching themes. To ensure accuracy, multiple responses from a single participant within a category or theme were counted once, reflecting the total number of participants contributing to each category or theme.

Table 6 summarizes the thematic analysis results, outlining 11 categories from which two overarching themes emerged: (1) Benefits of AI Text-to-Image Tools and (2) Challenges of AI Text-to-Image Tools. The category of 'Inconsistencies' has been merged into 'Limitations in AI Outputs' to streamline the thematic structure, as their meanings overlap significantly. The number of participants contributing to each category is noted, alongside sample responses.

Students widely reported the positive impacts of DALL-E on their vocabulary retention and learning experience. Visuals supported memory enhancement (11 participants), improving contextual understanding (12 participants). Additionally, students noted stronger word-image associations (16 participants), increased motivation (15 participants), more creative opportunities (9 participants), and the convenience (11 participants) of using DALL-E. Many appreciated learning AI as a skill (13 participants).

Despite the benefits, students faced challenges. Some noted limitations in AI output, including inconsistent or irrelevant visuals, which affected their learning (24 participants). Others had difficulty with skills like generating accurate prompts (13 participants). Distractions from visuals (9 participants) and cost issues (10 participants) were also mentioned.

Table 6

Thematic Analysis of Students' Perceptions of DALL-E in Vocabulary Retention

Samples of Responses (Participant #)	Codes	Categories (# of participants)	Themes
"The photos help me memorize the words better and longer." (S2)	Memorize better (6), retain longer (5)	**Memory Enhancement (11 participants):** DALL-E helps improve vocabulary recall and retention.	**Benefits of AI Text-to-Image Tools** Improves vocabulary retention, learning experience, creativity, and interaction.
"DALL-E visuals make remembering words easier because they stay in my mind." (S5)	Easier to remember (5), visuals stick in memory (4)		
"Using DALL-E increases motivation and makes the lesson more enjoyable." (S40)	Increases motivation (8), an enjoyable learning experience (7)	**Motivational Benefits (26 participants):** DALL-E makes learning enjoyable, increasing motivation to participate.	
"I loved how the images made the learning process fun and creative." (S35)	Learning is fun (5), fosters creativity (6)		
"The DALL-E pictures are so detailed that they help me understand the meaning of the words more easily." (S37)	Visuals clarify meaning (7), make learning easier (5)	**Contextual Understanding (12 participants):** Visuals enhance comprehension of word meanings in real-life contexts.	
"DALL-E helped me connect the words with the images in my head." (S25)	Creates word-image association (9), strengthens mental connections (7)	**Word-Image Association (16 participants):** DALL-E strengthens the association between vocabulary and visuals.	
"It helps me create stories with the words, making the vocabulary feel practical and useful." (S7)	Practical use of words (6), vocabulary feels real (5)	**Practical Vocabulary Use (11 participants):** Using vocabulary in stories improves usability.	
"Using DALL-E for visual storytelling sparked my creativity and imagination." (S17)	Promotes creativity (4), fosters imagination (5)	**Creativity and Imagination (9 participants):** DALL-E promotes creativity in storytelling.	
"DALL-E saved me so much time because I did not have to draw or search for pictures manually." (S20)	Convenient (6), saves time (5)	**Convenience (11 participants):** DALL-E is easy to use and saves time when creating visuals.	
"Sometimes DALL-E generates images that do not match my story, which	Generates irrelevant images (7), mismatched visuals (6), inconsistent	**Limitations in AI Outputs (35 participants):**	**Challenges of AI Text-to-Image Tools**

is frustrating." (S10)	visuals (6), confusing characters (5)	DALL-E sometimes produces unclear, irrelevant, or inconsistent visuals.	- Technical limitations - Skill-based issues - Distraction - Costs
"The images can be unclear and hard to relate to the words." (S12)	Unclear visuals (6), hard to relate to vocabulary (5)		
"Inconsistent visuals for the same character make the story confusing" (S33)	Visuals are distracting (5), focus shifts away from vocabulary (6)		
"It's hard to figure out how to make the perfect prompts to get the images I want." (S22)	Difficulty creating prompts (8), lack of AI skills (5)	**Skill Challenges (13 participants):** Students face challenges in using DALL-E effectively.	
"I got distracted by the visuals and sometimes forgot to focus on the vocabulary." (S29)	Visuals are distracting (5), focus shifts away from vocabulary (4)	**Distraction Challenges (9 participants):** DALL-E visuals can sometimes reduce focus on language learning goals.	
"Some AI tools are expensive, and it's hard for students to afford them." (S18)	Costs too much (6), expensive for students (4)	**Costs (10 participants):** The high cost of DALL-E is a barrier for students.	

Discussion

This study explored how integrating DALL-E into digital storytelling supports vocabulary retention in English learners, revealing both benefits and limitations. Results showed that the DALL-E group outperformed the control group in both immediate and delayed vocabulary retention, aligning with Ahmad's (2019) findings that combining text and images enhances memory through dual cognitive pathways. It also supports Nematollahi et al. (2017), who highlighted the impact of visual storytelling on vocabulary learning. However, this study extends the literature by introducing DALL-E as an active tool for learners to generate personalized and contextually relevant visuals, aligning with Dual Coding Theory (Paivio, 1971). Unlike Ahmad's static multimedia glosses or Nematollahi's pre-made visuals, DALL-E enables learners to actively create content actively, enhancing engagement and retention. This active creation aligns with Constructivism, where learners construct knowledge through hands-on activities (Piaget, 1972).

Students reported that DALL-E increased their motivation, creativity, and contextual understanding of vocabulary. These findings support Zabidin's (2015) research on the role of emotional engagement in vocabulary retention and expand on Leong et al. (2019), who emphasized the motivational benefits of digital storytelling. Students appreciated the convenience and practicality of DALL-E, which supported Albishi and Alqiawi's (2022) conclusion on how DST boosts engagement. However, challenges such as inconsistent visuals, reliance on AI, and technical issues were noted. For instance, mismatched visuals disrupted

narrative coherence, potentially reducing vocabulary focus, as reported by 11 participants (Table 5). This aligns with Żelaszczyk and Mańdziuk (2024), who highlighted AI's limitations with complex prompts. Contextual factors, such as students' limited prior experience with AI tools (noted in the tutorial session) and the novelty of technology, may have contributed to these challenges. To address this, future studies could extend training sessions to improve prompt-crafting skills and explore open-source AI tools to mitigate cost barriers. Additionally, incorporating teacher feedback on AI-generated visuals could ensure alignment with learning objectives, reducing distractions.

This study adds to existing research by combining DALL-E with DST to create a novel multimodal approach for vocabulary retention. Unlike previous studies using static visuals, DALL-E's dynamic content offers a personalized learning experience that adapts to individual needs, supporting Multimodal Learning (Mayer, 2002). The results provide quantitative evidence of improved retention (Tables 3 and 4) and qualitative insights into learner perceptions (Table 5), confirming the efficacy of this approach. The study also strengthens the application of Krashen's Input Hypothesis by ensuring that DALL-E-generated visuals align with learners' i+1 levels, making vocabulary input comprehensible and engaging. However, the short intervention period limited the exploration of long-term effects, and future research should compare DALL-E with other AI tools to isolate its specific contributions.

Implications for Pedagogy

First, teachers should use DALL-E in lessons to make vocabulary learning more motivating and relevant. These tools help students retain words and learn key technological skills. However, students should use DALL-E wisely to preserve their creativity. Teachers can promote critical engagement by asking students to refine AI-generated visuals, ensuring active involvement and alignment with learning goals. AI can create images quickly, but too much dependence may limit creativity and problem-solving. Teachers should encourage students to question and adjust the images so that they stay active and the visuals serve the lesson. Clear instructions are also needed to manage technical issues and select reliable tools. Using open-source programs like Stable Diffusion can lower costs and improve access. Finally, combining text, images, and group tasks makes learning more effective. Group storytelling, in particular, helps students remember new words and practice communication.

Limitations of the Study and Recommendations for Future Action

This study has several limits. The sample was small, with only two classes, so the results may not reflect larger groups. It also focused on short-term outcomes, leaving the long-term impact of DALL-E on vocabulary learning uncertain. Student reports may contain bias, and prior technology experience was not considered.

For practice, some steps are suggested. Teachers need training to use DALL-E effectively and solve technical problems. Course design should add digital storytelling so students can use and evaluate AI images more critically. Future work should test longer programs to check lasting effects. Developers should improve image accuracy, while schools and policymakers can reduce costs by working with AI providers or adopting open-source tools.

Conclusion

This study found that combining DALL-E with digital storytelling (DST) helped second-year English majors in Vietnam retain vocabulary more effectively. The experimental group outperformed the control group in both short- and delayed tests. Students also noted stronger

word–image links, supporting Dual Coding Theory (Paivio, 1971) and Krashen's Input Hypothesis (1982).

Although issues such as uneven outputs and cost were noted, the six-week program showed that, with training and suitable tools, DALL-E can be used effectively in class. Its creative use reflects constructivist principles (Piaget, 1972), allowing students to build knowledge in a more personal and active way. This approach provides a practical alternative to rote learning in EFL classrooms. Further research should explore the long-term impact and test the model with other learner groups.

Acknowledgments

This work was funded by Van Hien University and Van Lang University. The authors sincerely thank both institutions for their support, which made the study possible.

References

Ahmad, S. Z. (2019). Multimedia glosses for enhancing EFL students' vocabulary acquisition and retention. *English Language Teaching, 12*(12), 46–58. https://doi.org/10.5539/elt.v12n12p46

Albishi, O. A., & Alqiawi, D. A. (2022). The role of digital storytelling in the improvement of vocabulary acquisition. *International Journal of English Language Studies, 4*(1), 132–142. https://doi.org/10.32996/ijels.2022.4.1.10

Allanazarova, M. (2020). Vocabulary retention in cognitive theory. *Bulletin of Science and Practice, 6*(9), 414–420. https://doi.org/10.33619/2414-2948/58

Al-Obaydi, L. H., & Pikhart, M. (2024). Revisiting Total Physical Response: Evaluating its impact on vocabulary acquisition and retention in EFL classrooms. *Forum for Linguistic Studies, 6*(5), 822–832. https://doi.org/10.30564/fls.v6i5.7028

Alsadoon, R. (2021). Chatting with AI bot: Vocabulary learning assistant for Saudi EFL learners. *English Language Teaching, 14*(6), 135–149. https://doi.org/10.5539/elt.v14n6p135

Ashkan, L., & Seyyedrezaei, S. H. (2016). The effect of corpus-based language teaching on Iranian EFL learners' vocabulary learning and retention. *International Journal of English Linguistics, 6*(4), 190–200. https://doi.org/10.5539/ijel.v6n4p190

Baddeley, A. D. (1997). *Human memory: Theory and practice*. Psychology Press.

Baisel, D. A., & Ramachandran, S. (2024). Fostering vocabulary memorization: Exploring the impact of AI-generated mnemonic keywords on vocabulary learning through Anki flashcards. *World Journal of English Language, 14*(2), 434–446. https://doi.org/10.5430/wjel.v14n2p434

Barua, S. (2023). Digital storytelling: Impact on learner engagement and language learning outcomes. *International Journal of Academic and Applied Research, 7*(6), 25–39. https://www.researchgate.net/publication/372110349_Digital_Storytelling_Impact_on_Learner_Engagement_and_Language_Learning_Outcomes

Belda-Medina, J., & Goddard, M. B. (2024). The effect of digital storytelling on English vocabulary learning in inclusive and diverse education. *International Journal of English*

Language Studies, 6(1), 110–118. https://al-kindipublisher.com/index.php/ijels/article/view/6869

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

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

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

Hill, A. C. (2022). The effectiveness of mnemonic devices for ESL vocabulary retention. *English Language Teaching, 15*(4), 6–15. https://doi.org/10.5539/elt.v15n4p6

Jamal, S., Wimmer, H., & Rebman, C. M. (2024). Perception and evaluation of text-to-image generative AI models: A comparative study of DALL-E, Google Imagen, GROK, and Stable Diffusion. *Issues in Information Systems, 25*(2), 123–134. https://doi.org/10.48009/2_iis_2024_123

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Le, V. H. H. (2020). Digital storytelling with puppet pals to generate freshmen's enjoyment in English speaking. *Computer-Assisted Language Learning Electronic Journal, 21*(3), 175–197. https://callej.org/index.php/journal/article/view/318/249

Le, V. H. H., & Doan, O. T. K. (2022). Comic books: Overcoming challenges in online collaborative learning. *International Journal of Computer-Assisted Language Learning and Teaching, 12*(4), 1–22.

Leong, A. C. H., Abidin, M. J. Z., & Saibon, J. (2019). Learners' perceptions of the impact of using digital storytelling on vocabulary learning. *Teaching English with Technology, 19*(4), 3–26. https://files.eric.ed.gov/fulltext/EJ1233478.pdf

Lim, N. Z. L., Zakaria, A., & Aryadoust, V. (2022). A systematic review of digital storytelling in language learning in adolescents and adults. *Education and Information Technologies, 27*(5), 6125–6155. https://doi.org/10.1007/s10639-021-10861-0

Long, N. T., & Van, V. H. (2019). Change method of teaching and learning at universities in Vietnam on current. *International Journal of Advance Research, Ideas and Innovations in Technology, 5*(4), 413–418. https://www.ijariit.com/manuscripts/v5i4/V5I4-1300.pdf

Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85–139). Academic Press.

Moelyono, T. P. A., Murtisari, E. T., Kurniawan, D., & Thren, A. (2023). Google Translate in EFL freshmen's writing assignments: Uses, awareness of benefits and drawbacks, and perceived reliance. *Vision: Journal for Language and Foreign Language Learning, 12*(1), 47–66. https://doi.org/10.21580/vjv12i217446

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Nematollahi, B., Behjat, F., & Kargar, A. A. (2017). The effect of aural and visual storytelling on vocabulary retention of Iranian EFL learners. *English Language Teaching, 10*(4), 92–104. https://doi.org/10.5539/elt.v10n4p92

Nguyen, T. A., & Jaspaert, K. (2022). Implementing task-based language teaching in an Asian context: Is it a real possibility or a nightmare? *ITL - International Journal of Applied Linguistics, 173*(1), 31–50. https://doi.org/10.1075/itl.16022.ngu

Paivio, A. (1971). *Imagery and verbal processes*. Psychological Press.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Piaget, J. (1972). *Psychology and epistemology: Towards a theory of knowledge*. Allen Lane.

Pratt, Y. P. (2019). *Digital storytelling in Indigenous education: A decolonizing journey for a Métis community*. Routledge.

Robin, B. R. (2008). Digital storytelling: A powerful technology tool for the 21st-century classroom. *Theory Into Practice, 47*(3), 220–228. https://doi.org/10.1080/00405840802153916

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Sembiring, D. L. B., & Simajuntak, D. C. (2023). Digital storytelling as an alternative teaching technique to develop vocabulary knowledge of EFL learners. *Journal of Languages and Language Teaching, 11*(2), 211–224.

Sweller, J. (1988). Cognitive load during problem-solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Tran, T. T. (2013). The causes of passiveness in learning of Vietnamese students. *VNU Journal of Science: Education Research, 29*(2), 72–84.

Vu, D. V., & Peters, E. (2021). Vocabulary in English language learning, teaching, and testing in Vietnam: A review. *Education Sciences, 11*(9), 1–11.https://awspntest.apa.org/record/2023-25435-002

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

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(1), 33–52. https://doi.org/10.1017/S0272263105050023

Yang, W., & Dai, W. (2011). Rote memorization of vocabulary and vocabulary development. *English Language Teaching, 4*(4), 61–64. http://dx.doi.org/10.5539/elt.v4n4p61

Zabidin, N. B. (2015). The use of humorous texts in improving ESL learners' vocabulary comprehension and retention. *English Language Teaching, 8*(9), 104–115. https://doi.org/10.5539/elt.v8n9p104

Żelaszczyk, M., & Mańdziuk, J. (2024). Text-to-image cross-modal generation: A systematic review. *arXiv:2401.11631*. 1–148. https://arxiv.org/pdf/2401.11631

Zhelyazkov, Y. A. (2024). AI assistance in language education: AI-detection accuracy and students' vocabulary retention. *Journal for Research Scholars and Professionals of English Language Teaching, 8*(44). 1–8. https://doi.org/10.54850/jrspelt.8.44.001

Biodata

Le Thi Kieu Van, PhD, is a senior lecturer and Dean of the Faculty of Foreign Languages at Van Hien University, Vietnam. Her research interests focus on developing language skills, syllabus design, especially ESP courses, and cognitive linguistics. She has already published textbooks and articles in applied linguistics and has also presented at various international conferences.

Van Huynh Ha Le (M.A) is an English lecturer at Van Lang University, Vietnam, with degrees from the University of Pedagogy (HCMC) and Victoria University (Australia). She has presented at various international conferences and is pursuing a PhD at Burapha University, Thailand. Her research focuses on using technology to boost student motivation in English learning. ORCID: 0000-0001-8473-5351.