Technology-enhanced vocabulary learning: the role of self-regulation and prior knowledge

Identification Number/DOI: 10.64152/10125/73640 <https://doi.org/10.64152/10125%2F73640>

Publisher: National Foreign Language Resource Center

Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND
2025, Volume 29, Issue 1
pp. 1–31

ARTICLE

-

Technology-enhanced vocabulary learning: The role of self-regulation and prior knowledge

Jiarun Ye, University of Reading
Pengchong Zhang, University of Reading
Rowena Kasprowicz, University of Reading
Catherine Tissot, University of Reading

Abstract

Technology-enhanced vocabulary learning has gained increasing attention, yet its effectiveness remains inconclusive, with studies reporting mixed findings. Additionally, whether integrating self-regulated learning mechanism (SLM) into such type of learning can further enhance its effectiveness remains underexplored. This study examines the impact of two approaches, digital flashcards (DF) and video enhancement (VE), with and without SLM on vocabulary learning. It also investigates how pre-existing vocabulary knowledge (PVK) and self-regulation moderate these effects. Conducted with 132 junior high Chinese EFL learners over six weeks, pre- and post-tests measured written and aural form recognition and meaning recall across four experimental conditions and a control group. Results indicated that all interventions significantly improved vocabulary learning, with VE outperforming DF. The SLM was effective for supporting form recognition, particularly in VE, but its effectiveness diminished for the more demanding task of meaning recall. Self-regulation did not moderate learning outcomes, but PVK did for meaning recall within DF. Higher PVK learners benefited more from additional SLM, as it allowed them to integrate prior knowledge, whereas learners with limited PVK preferred DF alone. These findings highlight the need to align technology-enhanced vocabulary learning with task complexity and individual differences, offering practical insights for more adaptive instructional design.

Keywords: technology-enhanced vocabulary learning; digital flashcards, video enhancement; self-regulated learning mechanism

Language(s) Learned in This Study: *English*

APA Citation: Ye, J., Zhang, P., Kasprowicz, R., & Tissot, C. (2025). Technology-enhanced vocabulary learning: The role of self-regulation and prior knowledge. *Language Learning & Technology*, *29*(1), 1–31. https://doi.org/10.64152/10125/73640

Introduction

Mastering vocabulary knowledge is essential for second language (L2) proficiency, as it strongly predicts overall language ability (Nation, 2022). However, vocabulary learning is challenging since knowing a word involves form, meaning, and usage, both receptively (e.g., recognizing word forms) and productively (e.g., recalling meanings), with the latter being more demanding (Nation, 2022). Vocabulary learning can occur incidentally, for example through reading or listening, or intentionally through targeted learning activities (Webb & Nation, 2017). Studies suggest intentional learning leads to better short-term outcomes (Webb, 2005), while both methods show similar, though limited, long-term retention (Schmitt, 2008). Over the past two decades, technology-enhanced vocabulary learning has emerged, integrating multimodality (Mayer, 2005) to improve both incidental and intentional learning, particularly for productive vocabulary retention.

*** Corresponding Author:** Pengchong Zhang, anthony.zhang@reading.ac.uk

Technology-enhanced vocabulary learning involves using digital tools (e.g., mobile apps, computers) and resources (e.g., videos, images) to support vocabulary learning (Zhou et al., 2024). Two common methods are digital flashcards (DF) and video enhancement (VE). DF promote retrieval practice (Laufer et al., 2004) and feedback (Nation, 2022) through interactive tasks like multiple-choice questions (Li et al., 2021). In contrast, VE provides contextualized vocabulary exposure through stories and dialogues, engaging learners via multiple modalities (Mayer, 2005). Despite reported benefits, many empirical studies on DF and VE for vocabulary learning (e.g., Ghoorchaei et al., 2021; Yüksel et al., 2020) have not included a control group to account for baseline effects. Moreover, most of these studies have focused on adult learners, which limits the generalizability of their findings to younger populations. In addition, few studies have directly compared DF and VE. The only known study (Bueno-Alastuey & Nemeth, 2020) found no significant differences between DF and VE. Further research is therefore needed to clarify the relative effectiveness of these approaches, particularly with younger learners.

Another key area of study is whether integrating a self-regulated learning mechanism (SLM) into technology-enhanced vocabulary learning can further improve learning outcomes. A SLM involves goal-setting, note-taking, and evaluation modules, which have been shown to enhance intrinsic motivation (Chen et al., 2019). This is particularly important for independent learning, as it enhances sustained engagement and perseverance (Wang et al., 2020). Recent research has explored SLM's role in rote learning of wordlists containing with L1 translations and example sentences (Chen et al., 2019; Wang et al., 2020). It remains unclear, however, whether the inclusion of SLM would benefit more engaging approaches, such as DF and VE, due to the absence of studies that have combined SLM with either approach. Specifically, further research is needed to determine if SLM enhances learning by promoting active engagement, self-directed exploration, and deeper vocabulary processing.

One more object of study is the moderating role of individual learner differences in technology-enhanced vocabulary learning, as learners with varying prior knowledge and learning backgrounds may respond differently to instructional tasks (Paas et al., 2003). Within vocabulary learning, pre-existing vocabulary knowledge (PVK) plays a crucial role. Research shows that higher PVK learners benefit more from complex tasks (e.g., L2 explanations, contextual clues), whereas lower PVK learners perform better with simpler, structured tasks (e.g., L1-support instruction) (El-Dakhs et al., 2022; Zhang & Graham, 2020). Additionally, self-regulation is particularly important for lower PVK learners, helping them manage and process task-related information more effectively (Deng & Trainin, 2023). Despite these findings, few studies have examined how PVK and self-regulation moderate learning outcomes in technology-enhanced vocabulary learning.

In summary, the current study (a) examines whether redesigning and comparing DF and VE on a unified digital platform provides a clearer understanding of their differences in supporting vocabulary learning; (b) assesses whether integrating an additional SLM with DF and VE can further enhance vocabulary learning; and, (c) understands how learners' PVK and self-regulation influence vocabulary learning outcomes. These research gaps are discussed in detail in the Background Literature section.

Background Literature

This study investigates the effectiveness of DF and VE in vocabulary learning, informed by a range of theoretical perspectives relevant to task design, learner differences, and multimodality. These include the Involvement Load Hypothesis (ILH) (Laufer & Hulstijn, 2001), models of self-regulated learning (Zimmerman & Schunk, 2001), and cognitive theories such as Dual Coding (Paivio, 1986), Cognitive Load (Paas et al., 2003), and multimedia learning (Mayer, 2005). These perspectives underpin both the development of the instructional conditions and the interpretation of learner outcomes. Their relevance to DF, VE, and SLM is addressed throughout the following review of existing research.

Digital Flashcards

DF, a widely used technology-enhanced vocabulary learning approach, works like an electronic version of paper flashcards for learning and memorization (Li et al., 2021). Unlike traditional flashcards, DF are more flexible, allowing access via smartphones or computers. Apps such as Quizlet improve the experience by adding multiple-choice questions and giving instant feedback when learners make mistakes. These tasks support retrieval practice (Laufer et al., 2004) by repeatedly asking learners to recall and recognize word meanings, which helps vocabulary retention. Instant feedback also helps learners spot and correct errors, reinforcing memory (Nation, 2022), making DF effective for vocabulary learning.

Empirical studies have demonstrated the effectiveness of DF in L2 vocabulary learning. Yüksel et al. (2020) compared DF to wordlists in a within-subject study with 57 pharmacy students. They found that DF were more effective for technical vocabulary learning. Guessing when completing multiple-choice tests, however, may affected the reliability of the results. Therefore, Li and Hafner (2021) expanded previous study by comparing digital and paper flashcards among 85 undergraduates. They used a broader range of assessments for receptive and productive vocabulary knowledge. Their findings demonstrated the overall advantages of DF. Although these studies highlight the benefits of DF, they share limitations. Most lacked control groups, making it difficult to attribute learning gains solely to DF. Additionally, most research has focused on university students. This leaves a gap in understanding DF's effects on younger learners, despite the crucial role for early vocabulary learning in language development and academic success (Webb & Nation, 2017).

The potential benefits of DF can be explained through the ILH (Laufer & Hulstijn, 2001), a framework for designing effective vocabulary learning tasks. ILH builds on Craik and Tulving's (1975) Depth of Processing theory, identifying three key components for vocabulary retention: *need*, *search*, and *evaluation*. *Need*, learner's motivation, which is moderate when externally driven (e.g., teacher instruction) and strong when intrinsically motivated. *Search* refers to actively finding a word's meaning or form, with stronger *search* requiring independent retrieval. *Evaluation* means comparing or assessing word meanings, with stronger *evaluation* involving contextual understanding. Tasks that involve *need*, *search*, and *evaluation* promote deeper processing, leading to better vocabulary retention (Laufer & Hulstijn, 2001). According to ILH, DF trigger a strong level of *search* since learners must retrieve and recall word meanings rather than receiving them passively. However, the levels of *need* and *evaluation* remain moderate.

Video Enhancement

In contrast to DF, another technology-enhanced vocabulary learning approach, VE, engages a different aspect of the ILH, *evaluation*. VE provides authentic contexts (e.g., video, podcasts) that help learners understand how target words function in real-life language use (Ghoorchaei et al., 2021). It supports vocabulary learning by embedding words in meaningful linguistic contexts and combining visual and auditory elements to engage learners through multiple modalities. This approach aligns with dual coding theory (Paivio, 1986) and the cognitive theory of multimedia learning (Mayer, 2005), which suggest that processing multimodal input activates both verbal and non-verbal memory channels, increasing memory capacity and retention.

Several studies have demonstrated the positive impact of VE on vocabulary learning. Mashhadi et al. (2016) examined a blended learning module with 447 university students across three groups: self-study without podcasts, podcast-mediated blended learning, and face-to-face instruction. The podcast group, which used video podcasts featuring dialogues and stories with target vocabulary, showed the greatest vocabulary gains. Similarly, Ghoorchaei et al. (2021) studied podcast-based instruction with 60 university students. They found that the experimental group, which received 45-minute podcast sessions on pronunciation, meanings, and usage, significantly outperformed the control group in receptive vocabulary knowledge. Although research supports VE as effective tools for providing authentic language contexts, studies in this area remain limited. Most focus on adult learners, leaving a gap in understanding their effects on younger learners. Additionally, reliance on multiple-choice tests raises concerns about guessing and does not fully capture deep vocabulary retention and usage (Nation, 2022).

According to the ILH, the positive impact of VE may stem from its ability to trigger a high level of *evaluation*, though *need* and *search* remain moderate. By providing contextualized information, VE allows learners to compare target words with their linguistic context, promoting deeper word processing. DF and VE differ in their involvement load factors: DF primarily emphasizes *search*, and VE engages a strong level of *evaluation*. Recent research has debated the relative importance of *evaluation* within ILH. Yanagisawa and Webb (2022), in a meta-analysis, argue that *evaluation* may have a stronger impact on vocabulary gains because it requires learners to focus on both word form and syntagmatic use. In contrast, *need* primarily drives motivation, and *search* focuses on meaning retrieval, with each addressing only one aspect of learning. Consequently, tasks involving strong *evaluation* are more likely to trigger more in-depth vocabulary learning.

Based on Yanagisawa and Webb's (2022) findings, it can be hypothesized that VE with stronger *evaluation*, may be more beneficial than DF. However, the only known study directly comparing DF and VE (Bueno-Alastuey & Nemeth, 2020) among 23 adult learners (aged 18–61) found no significant difference in their effects on vocabulary retention (both receptive and productive knowledge). In that study, DF were delivered through Quizlet, allowing students to choose flashcard modes (e.g., multiple-choice, or fill-in-the-blank) for vocabulary practice. VE was implemented via Podcasts, providing an authentic linguistic context. In the Podcast group, students received word forms, meanings, and contextual materials in text format and were required to create and study from self-recorded podcasts. It should be noted that the use of different digital platforms might impact the reliability of the results, that is, Quizlet for DF, Podcasts for VE, as variations in interface design could have influenced learning outcomes (Kim & Kim, 2012). Therefore, more research is needed to compare DF and VE within the same platform and to determine whether stronger *evaluation* offers additional learning advantages.

Self-Regulated Learning Mechanism in Technology-Enhanced Vocabulary Learning

One important consideration when evaluating DF and VE using the ILH is that neither approach triggers a strong level of *need*, as learners primarily follow external instructions rather than engaging at their own pace. Yanagisawa and Webb (2022) highlighted the limited research attention given to the role of *need* in vocabulary learning. Even Laufer and Hulstijn (2001) acknowledged that *need* remains vaguely defined and is difficult to measure across different tasks. This suggests a need for further theoretical support on how to enhance learner motivation (Yanagisawa & Webb, 2022). In technology-enhanced vocabulary learning, one potential way to increase intrinsic motivation is through self-regulated learning (Chen et al., 2019). This is particularly relevant in the Chinese EFL learning context, where students are often driven by external motivation, such as exam-oriented goals (Zou et al., 2019). Therefore, enhancing intrinsic motivation through self-regulated learning may be especially beneficial in improving engagement and learning outcomes.

Self-regulated learning is when learners manage their motivation, emotions, behavior, cognition, and metacognition to reach their goals (Pintrich, 1995). It involves setting goals, tracking progress, and evaluating performance (Zimmerman & Schunk, 2001). This supports intrinsic motivation and links with the ILH by promoting a strong level of *need*. Many models have been developed to explore self-regulated learning, each focusing on different aspects. To bring these together, Puustinen and Pulkkinen (2001) identified three main phases: the preparatory phase (goal-setting and previewing), the performance phase (monitoring progress), and the appraisal phase (reviewing and adjusting).

Building on the self-regulated learning process, Chen et al. (2019) developed an English vocabulary learning app with an integrated self-regulated learning mechanism (EVLAPP-SLM), including goal-setting, note-taking, and evaluation modules. A study with 46 fifth-grade EFL learners found that those using EVLAPP-SLM outperformed the control group in both receptive and productive vocabulary knowledge. Among adult learners, Wang et al. (2020) introduced a collaborative app, Contribution-oriented Self-Directed Mobile Learning Ecology (CSDMLE), involving 55 university freshmen. This model emphasized peer learning and active recall, using WeChat groups to share vocabulary illustrations and feedback. While the CSDMLE group showed better immediate post-test performance (meaning recognition tests), no long-term retention advantage was observed. Both studies suggest that SLM enabled learners to take an active role, enhancing intrinsic motivation (*need* in ILH) and improving vocabulary learning. These studies, however, relied on rote learning methods, which offer limited interaction and engagement. Thus, incorporating SLM into DF and VE might yield better learning outcomes. Moreover, no study has yet explored whether integrating SLM into DF or VE would produce significant advantages compared to using DF or VE alone, highlighting a key research gap.

Individual Differences in Technology-Enhanced Vocabulary Learning

Individual differences, especially PVK, are key in vocabulary learning, affecting how learners handle tasks. Research shows that learners with higher PVK benefit more from complex designs like contextual inference, but those with lower PVK do better with simpler, L1-supported instruction. El-Dakhs et al. (2022) found that learners with strong PVK used context effectively to learn phrasal verbs, but lower PVK learners struggled with extra L2 input. Zhang and Graham (2020) also showed that lower PVK learners found L2-only instruction difficult but gained more from L1 support, such as code-switching. These findings support Cognitive Load theory (Paas et al., 2003), which suggests complex designs increase extraneous load for lower PVK learners, making learning harder. Managing intrinsic and germane load, while reducing extraneous load, can improve learning for these learners.

Pre-existing self-regulation is another key factor that may moderate learning effectiveness, particularly when SLM are integrated. Deng and Trainin (2023) found that L2 learners with lower PVK who received self-regulation strategy interventions achieved significantly greater vocabulary gains than those with higher PVK. This suggests that as learners become more self-regulated, they are better able to manage extraneous cognitive load and engage with more complex instructional tasks effectively. Despite these insights, limited research has explored how PVK and self-regulation interact in technology-enhanced vocabulary learning. This highlights the need for further studies to examine how PVK and self-regulation together influence learning outcomes in digital learning environments.

Research Questions

Considering the above research gaps, the present study seeks to address the following research questions:

RQ1. To what extent do (a) digital flashcards and (b) video enhancement, with and without integrated self-regulated learning mechanism, differ in their effects on vocabulary learning?

RQ2. To what extent do learners' pre-existing vocabulary knowledge and self-regulation moderate the effects of the technology-enhanced vocabulary learning approaches?

Method

Participants

Participants included 132 Grade Eight Chinese EFL learners (aged 13–14) from three intact classes at two junior high schools in southern China. All participants voluntarily took part and had achieved Level 3 English proficiency under the Compulsory Education English Course Standards (MoE, 2017), equivalent to A2 in the Common European Framework of Reference for Languages. Before the study, informed consent was obtained from parents, teachers, and school headmasters, followed by child assent from each participant. In this quasi-experiment, the three classes were randomly assigned to three groups: DF (n = 50), VE (n = 50), and a control group (n = 32). To examine the impact of SLM, each experimental group was further divided into two conditions (with and without SLM, see Experimental Design). This resulted in five conditions: DF (n = 25), DFSLM (n = 25), VE (n = 25), VESLM (n = 25), and Control (n = 32).

Instruments

Target Vocabulary

Sixty target words were selected from the Grade Ten syllabus of the Compulsory Education English Course Standards (MoE, 2017). The words were divided into four lists of fifteen and carefully screened to exclude members of the same word family, affixes, or roots, to avoid reducing cognitive load and skewing memory-related outcomes (Schmitt, 2008). A full list of target words is provided in Appendix A.

To determine word frequency, all target words were analyzed using the Lextutor Vocabulary Profiler (https://www.lextutor.ca/vp/eng/). Results showed that 6 words (10%) fell into K1, 10 (16.67%) into K2, and the remaining 44 words (73.34%) were either in the Academic Word List (AWL; 22 words, 36.67%) or Off-List (22 words, 36.67%). Compleat VocabProfiler provided further breakdown: 65% of the words fell within K-1 to K-3 levels, making them broadly suitable for intermediate learners. Although nearly 74% of the words were low frequency, all were drawn from the official syllabus, ensuring their pedagogical relevance. To minimize prior exposure and isolate the effects of the intervention, we deliberately avoided high-frequency words that learners might already have known through extracurricular study. Moreover, learners at this level are accustomed to cognitively demanding tasks such as memorizing long vocabulary lists (Zou et al., 2019). The instructional approach adopted in this study focused on intentional vocabulary learning with explicit guidance, which helped manage cognitive demands despite the relative difficulty of the items.

Digital Platform

As noted in the literature review, many previous studies lacked a standardized digital platform for comparing vocabulary learning approaches. As a result, differences in display and interface may have affected outcomes. Kim and Kim (2012) showed that students learned more vocabulary using larger screens, highlighting how display size can influence learning. To address this, the current study used Modao (https://modao.cc/), a web-based platform that allows users to customize learning activities, to design the interventions. It supports images, videos, and interactive tasks, making it adaptable to various teaching aims. Learning activities can be shared via a simple web link, and no installation is needed (see Appendix B for sample links). In this study, to control device variability, all interventions were conducted in a multimedia classroom, where all students accessed the platform on the same type of computers.

Experimental Design

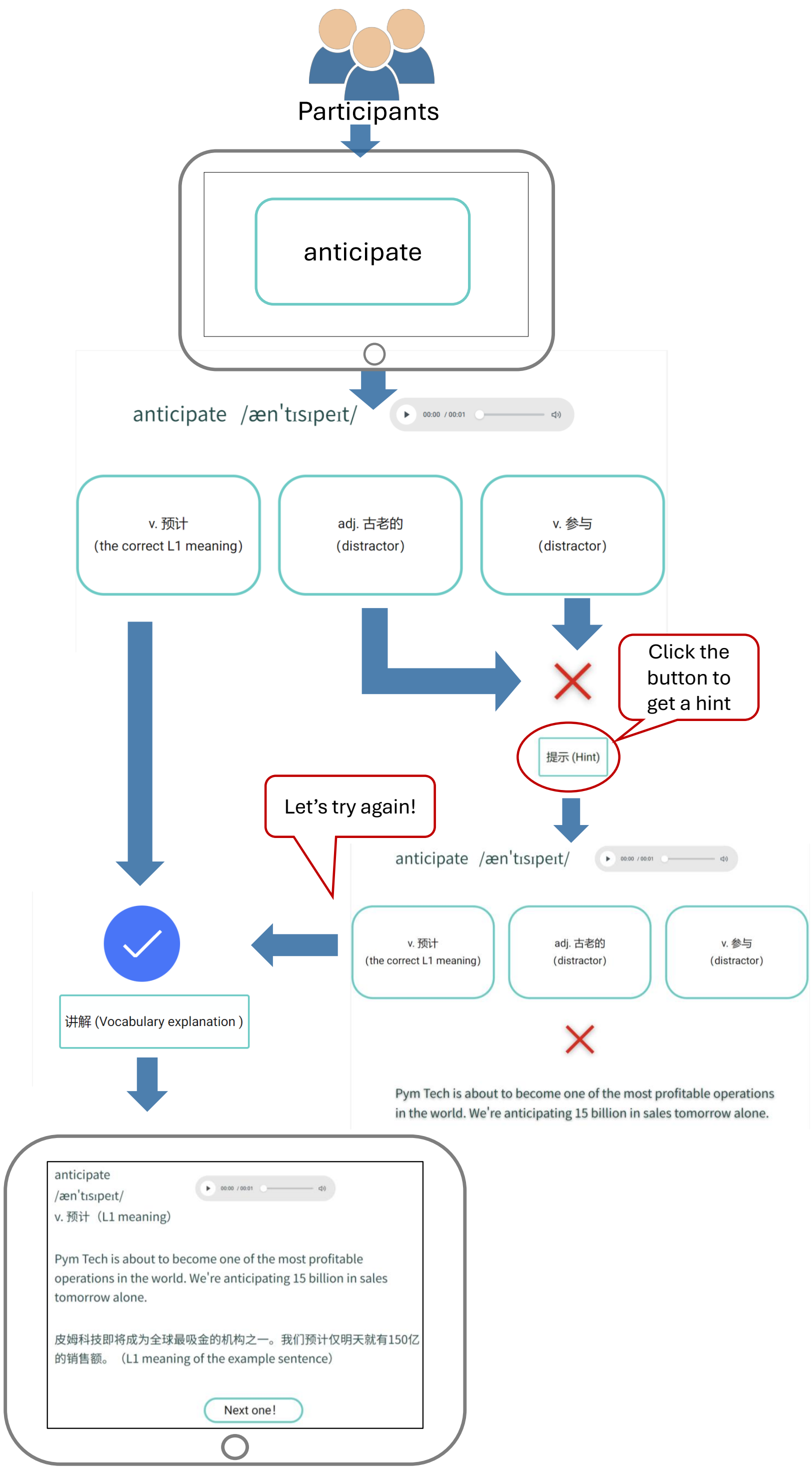
Following Nation's (2022) vocabulary knowledge framework, both experimental groups received vocabulary explanations covering the written form, pronunciation, Chinese translation, and an example sentence demonstrating usage. The key difference was the delivery approach—either via DF or VE. The DF included the target word's form, pronunciation, and three possible L1 meanings. If they answered incorrectly, a hint button appeared, revealing an example sentence for additional context. For correct responses, a link to the full vocabulary explanation page was provided. Figure 1 illustrates an example of DF.

Learners in DFSLM and VESLM received additional self-regulated learning (SLM) support, comprising three modules adapted from Chen et al. (2019): goal-setting, note-taking, and evaluation, shown in Figure 3. Before each session, the goal-setting module told participants how many vocabulary items to practice and let them set personal learning goals. During the session, the note-taking module allowed them to record their own strategies for each target word.

At the end of each session, the evaluation module assessed learners' progress through a vocabulary knowledge test. This also served as the pre- and immediate post-test for all participants. Unlike earlier studies relying on multiple-choice questions (e.g., Chen et al., 2019), this test followed Paribakht and Wesche's (1993) Vocabulary Knowledge Scale to assess both receptive and productive knowledge. Each word had three questions: (a) "Have you seen this word before?" (written form recognition); (b) "Have you heard this word before?" (aural form recognition); and (c) giving the meaning in Chinese or English (meaning recall). The first two items used yes/no responses, while the third required a written answer. All answers were scored as correct (1) or incorrect (0). The combination of receptive and productive measures gave a fuller view of vocabulary learning outcomes.

Figure 1

Example of DF



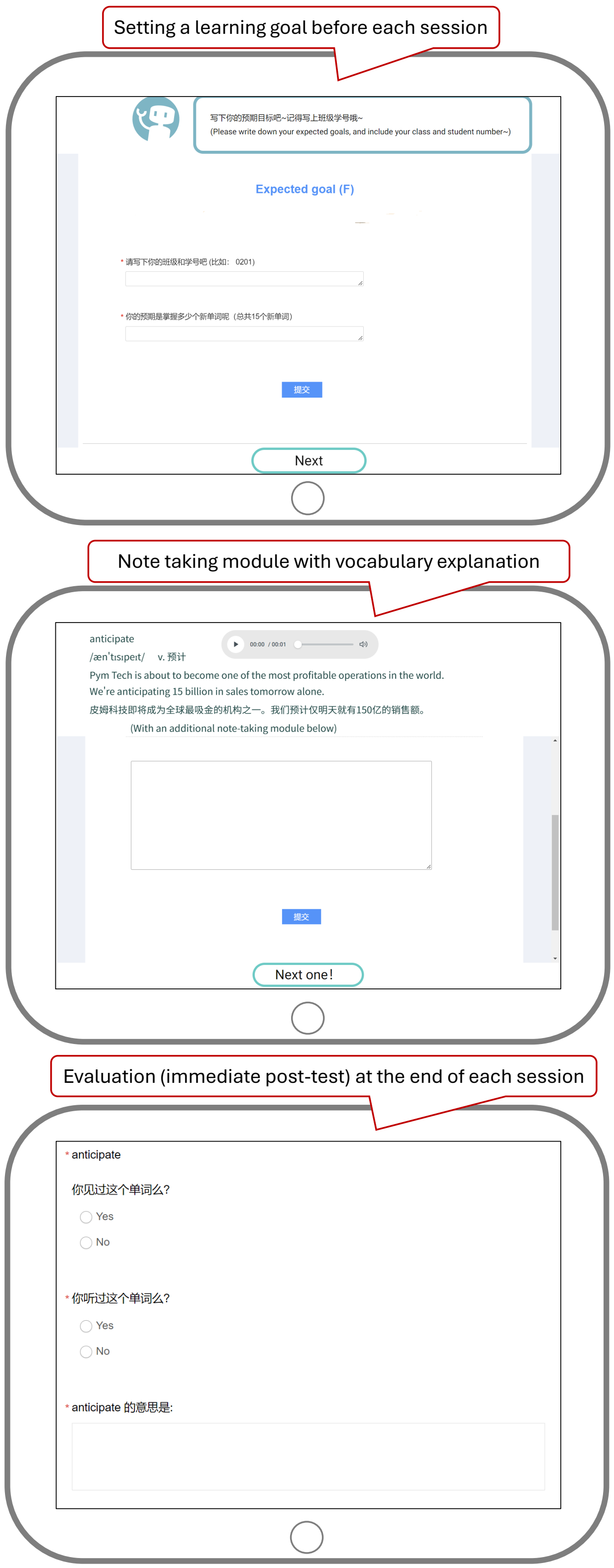
In VE, participants first accessed the vocabulary explanation page and then clicked a link button to watch a video clip extracted from a TV show, illustrating the target word in context. Figure 2 presents an example of VE.

Figure 2

Example of VE

Figure 3

Example of SLM



Measurements of Individual Differences

To assess participants' PVK before the intervention, a baseline vocabulary levels test was administered. Many online assessments, such as Meara and Miralpeix's (2016) yes/no word recognition format, may yield inaccurate results due to the potential for guessing. To address this, the study used the Vocabulary Levels Test (Version 2, Schmitt et al., 2001), which covers four levels (2000, 3000, 5000, and academic) with sixty words per level. As participants were junior high students with an estimated vocabulary of 1,600–2,000 words (Ministry of Education, 2017), the test could be too difficult. Therefore, word meanings were translated into Chinese, and the 5000-level and academic words were excluded. See Appendix C for an example.

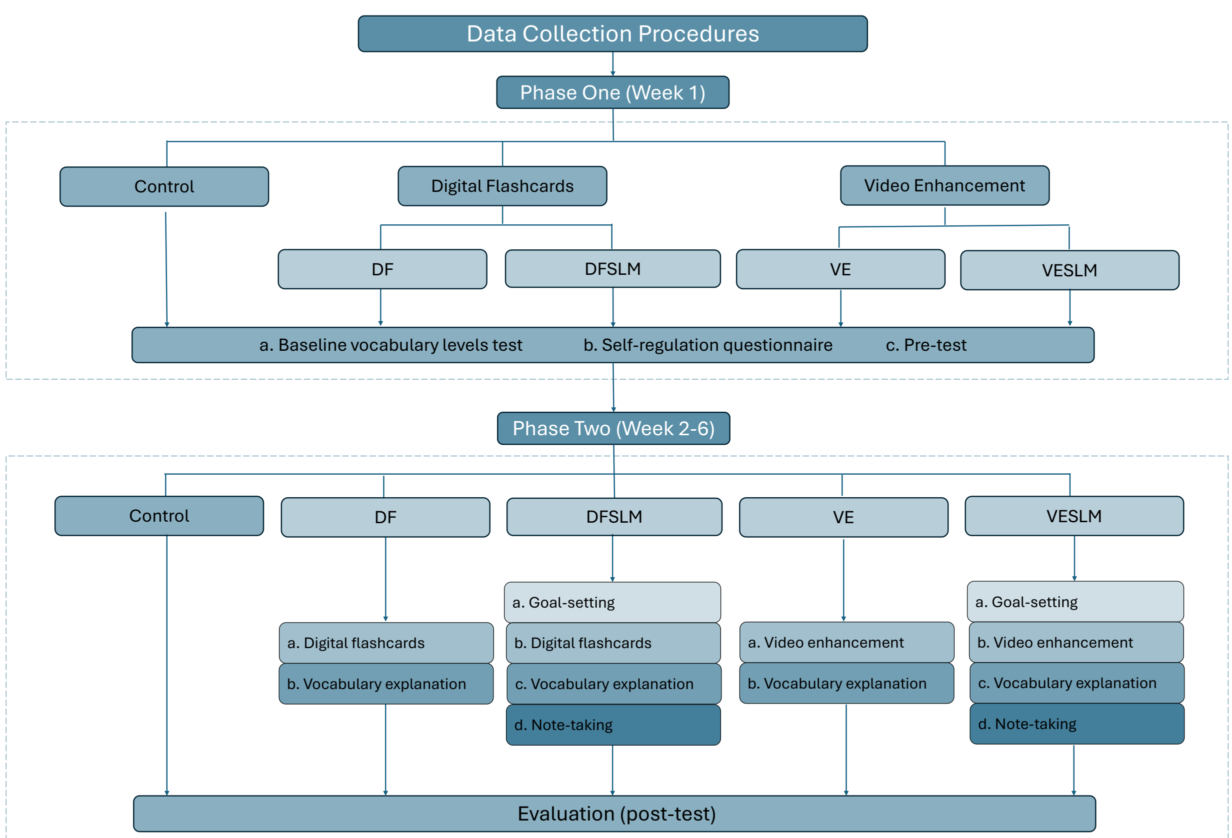
To measure self-regulation, the study used the questionnaire by Kızıl and Savran (2018), designed for technology-based vocabulary learning. It covers five factors: commitment (learning goals), metacognitive (focus management), affective (emotion regulation), resource (expanding resources), and social (seeking support). To fit this study's aims, the social factor was removed, as participants could not interact during the experiment to avoid cross-contamination. The term *information and communication technologies* was replaced with *vocabulary learning application (VLA)* for clarity. After revision, 17 items remained across four subscales: commitment (4), metacognitive (4), affective (6), and resource (3). See Appendix D for the revised questionnaire.

Procedures

Data collection procedures spanned six weeks across two phases, presented in Figure 4.

Figure 4

Data Collection Procedures



The study began with a pre-intervention phase where participants completed the baseline vocabulary levels test and self-regulation questionnaire to assess their PVK and self-regulation. A vocabulary pre-test also measured initial knowledge of the target words. The intervention phase took place in a school multimedia classroom, with each student using an individual computer. Over five weeks, the four experimental groups had one 45-minute session per week, each covering 15 target words. VESLM and DFSLM received additional SLM (goal-setting and note-taking), while VE and DF did not. All five groups completed the evaluation module as the immediate post-test. The control group completed the pre-intervention phase. They then continued with regular English lessons based on the national Grade Eight curriculum, using standard textbooks focused on general language skills. The target vocabulary was not covered, because the selected words came from the Grade Ten syllabus. As this group was from a different school, the risk of exposure to the intervention was minimal. After the experimental groups completed their final session, the control group took the vocabulary post-test.

Data Analysis

Reliability check was conducted among three measurement tools. Before the analysis, inter-rater reliability was assessed for meaning recall responses in the target vocabulary test. Since participants provided typed Chinese meanings, a second rater ensured scoring accuracy (Eisinga et al., 2013). A random sample of 3,960 items from both pre- and post-tests (out of 15,840) was analyzed. Cohen's kappa indicated high agreement: $\kappa = .95$ (95% *CI* [.93, .98]) for the pre-test and $\kappa = .96$ (95% *CI* [.94, .98]) for the post-test, exceeding the .80 benchmark (Fleiss et al., 2016). A further discussion was conducted regarding the disagreements and the final decision was applied consistently. Cronbach's alpha was used to evaluate the reliability (see Appendix E) and indicated that all measurements exhibited strong internal consistency, reaching the benchmark value (.80) proposed by Hinton et al. (2004).

To address the two research questions, data were analyzed by-Item and by-Participant using generalized linear mixed-effects models in R (R Core Team, 2024). Separate models were used to assess the effects of the interventions on written form recognition, aural form recognition, and meaning recall. Models were built with the *glmer()* function from the *lmerTest* package (Bates et al., 2015). The outcome variables were binary (0 or 1). Fixed factors included Time (pre- vs. post-test), Group (Control, DF, DFSLM, VE, VESLM), PVK_s (standardized *z*-scores), and SR_s (self-regulation, also standardized *z*-scores). Pre-test and Control served as baseline levels for Time and Group, respectively. The analysis had two steps: First, all models included Time, Group, their interaction (Time × Group), and PVK_s and SR_s as covariates. Then, if PVK_s or SR_s were significant, an extra model added three-way interactions (Time × Group × PVK_s or Time × Group × SR_s) to test for moderation effects in *RQ2*.

For each model, random effects were fitted stepwise to account for by-Participant and by-Item variations. Initially, only random intercepts were included, with fixed factors gradually added to the random slopes until further convergence was not possible. A binomial distribution was specified to match the binary outcome, and the *bobyqa* optimizer via the *glmerControl()* function ensured convergence in parameter estimation. Results were formatted in APA style using *tab_model()*, generating tables with coefficients and *p*-values. Interaction effects were visualized using the *ggplot2* package (Wickham, 2016), displaying predicted probabilities.

Results

Data supporting the results reported in this paper are openly available in Ye and Zhang (2025). Descriptive statistics were first calculated for all measurements (see Table 1). The results showed minimal vocabulary gains in the control group across the five-week period, despite continuing with regular English lessons based on the national curriculum. This offers a meaningful baseline for comparing technology-enhanced approaches.

Table 1

Descriptive Statistics

	Baseline vocabulary levels test			Self-regulation questionnaire		
Group	***M***	***SD***	**Range**	***M***	***SD***	**Range**
VE	1374.62	213.68	765.90 – 1764.90	58.88	12.79	37.00 – 90.00
VESLM	1414.59	235.15	956.70 – 1964.70	61.04	14.88	18.00 – 89.00
DF	1429.24	265.60	899.10 – 1998.00	62.72	11.74	36.00 – 90.00
DFSLM	1325.34	182.91	932.40 – 1731.60	65.32	12.28	39.00 – 90.00
Control	1085.37	595.70	266.40 – 998.00	76.56	10.16	54.00 – 90.00

Target Vocabulary Test

		Written form recognition			Aural form recognition			Meaning recall		
	Group	***M***	***SD***	**Range**	***M***	***SD***	**Range**	***M***	***SD***	**Range**
Pre-test	VE	29.40	10.51	11.00 – 50.00	27.20	10.06	9.00 – 50.00	10.28	7.52	1.00 – 33.00
	VESLM	29.12	10.95	4.00 – 53.00	25.52	10.17	5.00 – 53.00	10.40	9.55	0.00 – 45.00
	DF	29.12	12.14	12.00 – 59.00	25.56	13.12	6.00 – 59.00	13.60	14.41	0.00 – 50.00
	DFSLM	30.16	11.75	10.00 – 54.00	26.20	11.95	6.00 – 53.00	9.60	7.39	1.00 – 27.00
	Control	30.69	23.23	00.00 – 60.00	30.31	23.57	0.00 – 60.00	19.03	22.45	0.00 – 58.00
Post-test	VE	58.40	3.75	42.00 – 60.00	58.20	3.71	43.00 – 60.00	46.56	9.69	19.00 – 59.00
	VESLM	59.04	1.54	54.00 – 60.00	59.04	1.86	52.00 – 60.00	44.80	9.45	26.00 – 58.00
	DF	57.84	3.75	45.00 – 60.00	56.92	5.96	36.00 – 60.00	40.96	11.13	11.00 – 59.00
	DFSLM	58.72	3.66	43.00 – 60.00	56.88	7.40	26.00 – 60.00	37.08	9.84	16.00 – 54.00
	Control	31.69	23.23	0.00 – 60.00	30.81	24.00	0.00 – 60.00	20.81	22.15	0.00 – 59.00

Written Form Recognition

The final written form recognition model included Time, Group, and their interaction as fixed factors, with PVK_s and SR_s as covariates since they did not significantly predict the outcome variable. Results showed that all four treatment groups made significantly greater pre-post gains than the control group (see Table 2 and Figure 5). VESLM showed the highest improvement, with learners 110.34 times more likely to recognize written forms than those in control group. DFSLM (*OR* = 79.63) and VE (*OR* = 60.31) followed. Although DF (*OR* = 47.39) showed the least improvement, it still significantly outperformed the control group. Neither PVK nor self-regulation significantly predicted written form recognition.

Table 2

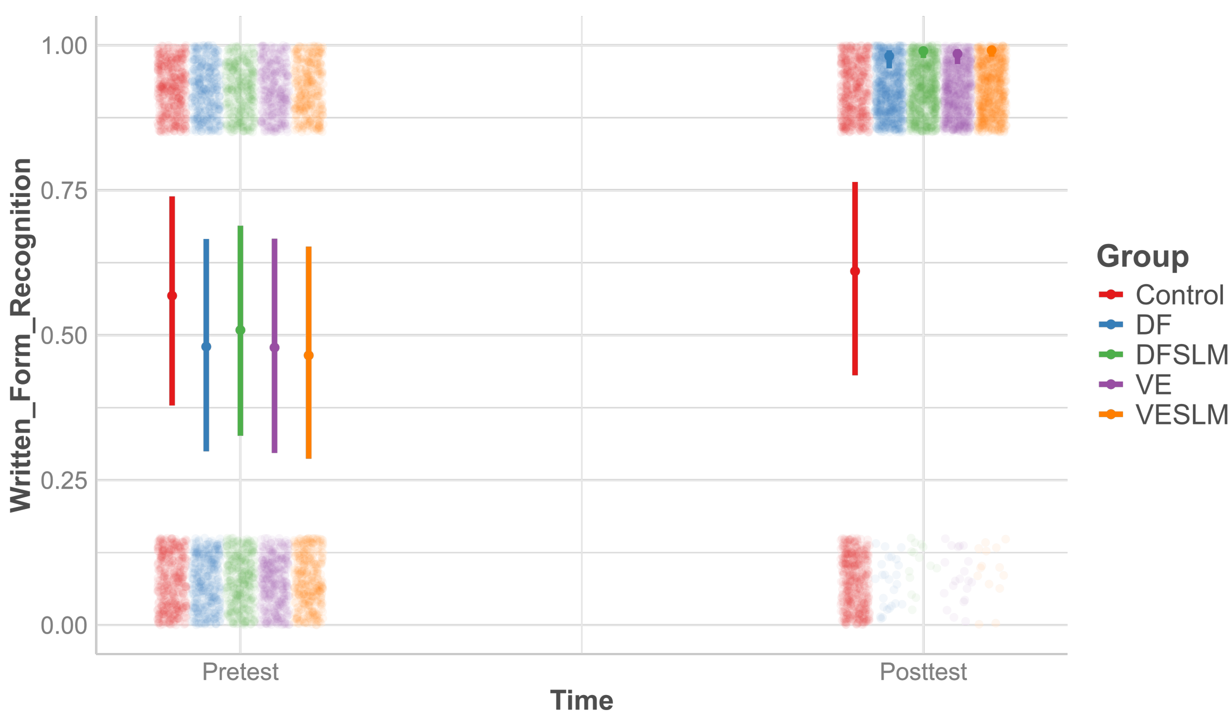
Results of Written Form Recognition Model

Predictors	***ORs***	**95% *CI***	***p***
(Intercept)	1.31	0.61 – 2.84	.486
Time $_{\text{Posttest-Pretest}}$	1.19	0.90 – 1.58	.221
Group $_{\text{DF-Control}}$	0.70	0.25 – 1.97	.501
Group $_{\text{DFSLM-Control}}$	0.79	0.29 – 2.12	.636
Group $_{\text{VE-Control}}$	0.70	0.24 – 2.00	.503
Group $_{\text{VESLM-Control}}$	0.66	0.23 – 1.87	.436
PVK_s	1.27	0.92 – 1.76	.151
SR_s	1.00	0.71 – 1.41	.999
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{DF-Control}}$	47.39	32.37 – 69.37	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{DFSLM-Control}}$	79.63	50.91 – 124.55	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{VE-Control}}$	60.31	39.92 – 91.14	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{VESLM-Control}}$	110.34	67.13 – 181.35	**< .001**
Observations	15840		
Marginal R^2 / Conditional R^2	0.383 / 0.722		

Note. *ORs* = odds ratios, *CI* = confidence interval.

Figure 5

Plot for Written Form Recognition Model



Aural Form Recognition

Similarly, the final aural form recognition model included only Time × Group interactions, as neither PVK_s nor SR_s significantly predicted the outcome. Results showed that all four treatment groups significantly outperformed the control group (see Table 3 and Figure 6). VESLM was again the most effective, with learners 138.00 times more likely to recognize aural forms than the control group. VE (*OR* = 59.89) and DF (*OR* = 46.45) followed, and DFSLM (*OR* = 42.32) was the least effective.

Table 3

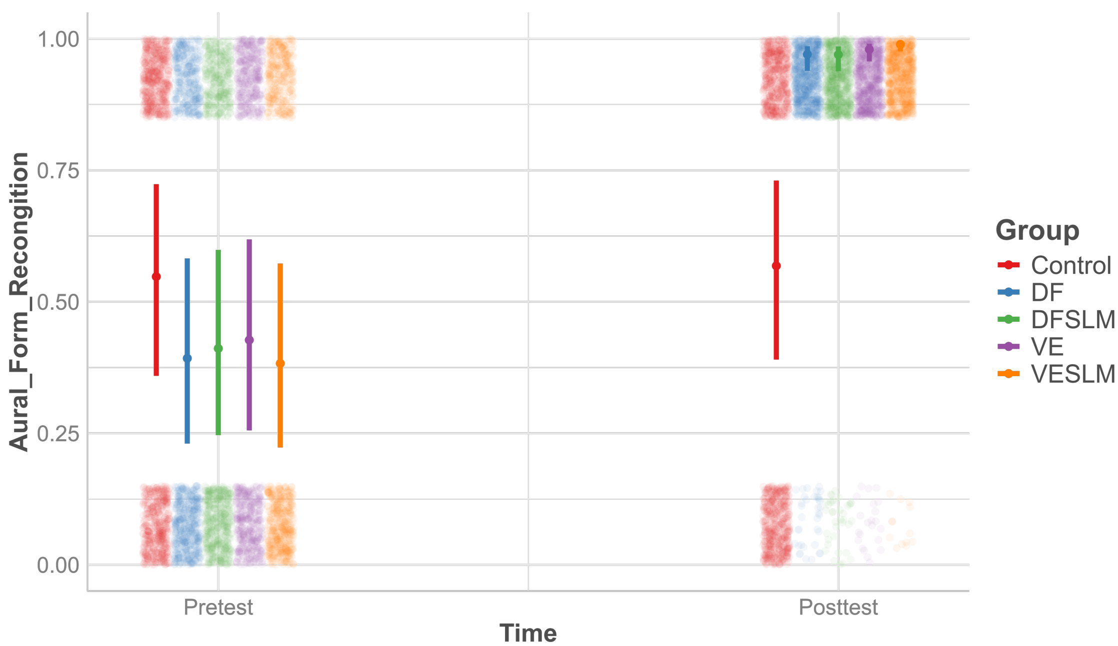
Results of Aural Form Recognition Model

	ORs	***CI***	***p***
(Intercept)	1.21	0.56 – 2.62	.626
Time $_{\text{Posttest-Pretest}}$	1.09	0.83 – 1.43	.551
Group $_{\text{DF-Control}}$	0.53	0.19 – 1.51	.235
Group $_{\text{DFSLM-Control}}$	0.58	0.21 – 1.56	.279
Group $_{\text{VE-Control}}$	0.62	0.21 – 1.77	.369
Group $_{\text{VESLM-Control}}$	0.51	0.18 – 1.46	.210
PVK_s	1.27	0.91 – 1.76	.155
SR_s	0.98	0.70 – 1.38	.912
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{DF-Control}}$	46.45	32.60 – 66.19	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{DFSLM-Control}}$	42.32	29.65 – 60.39	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{VE-Control}}$	59.89	40.34 – 88.90	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{VESLM-Control}}$	138.00	84.58 – 225.16	**< .001**
Observations	15840		
Marginal R^2 / Conditional R^2	0.361 / 0.707		

Notes. *ORs* = odds ratios, *CI* = confidence interval.

Figure 6

Plot for Aural Form Recognition



Meaning Recall

Unlike written and aural form recognition, the initial meaning recall model identified PVK as a significant covariate, but SR_s remained a non-significant covariate. A second model was built, adding three-way Time × Group × PVK_s interactions. Results (Table 4 and Figure 7) showed significant three-way interactions, indicating that learning gains varied by group depending on PVK levels. To clarify these moderation effects, pairwise comparisons were conducted using the *emmeans* package (Lenth, 2024), setting PVK at three proficiency levels: -2 (low PVK), 0 (PVK at Mean), and 2 (high PVK).

The pairwise comparison results (Table 5) showed the strongest moderation effects in the DF group, where PVK negatively influenced learning gains (*OR* = 2.57, 1/0.39). As PVK increased from -2 to 2, the odds of significant gains dropped from 155.22 to 4.63. The second-largest moderation effect was in the DFSLM group, but unlike DF, the effect was positive (*OR* = 2.05). Here, learning gains increased from 3.26 to 75.80 as PVK levels rose from -2 to 2. In the VESLM group, PVK negatively moderated learning gains (*OR* = 1.64, 1/0.61), with odds decreasing from 104.76 to 18.41 as PVK increased. Finally, no significant moderation effects of PVK were confirmed in the VE and control group.

After accounting for PVK's moderation effects, the results of the two-way interactions showed that all intervention groups made significant meaning recall gains (see Figure 8). The VESLM group had the highest odds ratio (29.11), making learners 29.11 times more likely to recall meanings than the control group. VE (*OR* = 27.92) and DF (*OR* = 17.77) followed, and DFSLM (*OR* = 10.42) showed the smallest improvement but still outperformed the control group.

Table 4

Results of Meaning Recall Model

	Ors	***CI***	***p***
(Intercept)	0.17	.08 – .36	**< .001**
Time $_{\text{Posttest-Pretest}}$	1.51	1.12 – 2.04	**.007**
Group $_{\text{DF-Control}}$	0.42	0.15 – 1.17	.097
Group $_{\text{DFSLM-Control}}$	0.67	0.26 – 1.73	.406
Group $_{\text{VE-Control}}$	0.69	0.25 – 1.91	.475
Group $_{\text{VESLM-Control}}$	0.44	0.15 – 1.24	.120
PVK_s	1.71	1.16 – 2.51	**.006**
SR_s	1.09	0.79 – 1.51	.609
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{DF-Control}}$	17.77	12.25 – 25.79	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{DFSLM-Control}}$	10.42	7.61 – 14.27	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{VE-Control}}$	27.92	19.94 – 39.11	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{VESLM-Control}}$	29.11	20.42 – 41.48	**< .001**
Time $_{\text{Posttest-Pretest}}$ × PVK_s	1.07	0.93 – 1.23	.346
Group $_{\text{DF-Control}}$ × PVK_s	5.53	1.96 – 15.60	**.001**
Group $_{\text{DFSLM-Control}}$ × PVK_s	0.67	0.16 – 2.79	.584
Group $_{\text{VE-Control}}$ × PVK_s	0.71	0.21 – 2.40	.582
Group $_{\text{VESLM-Control}}$ × PVK_s	2.54	0.81 – 7.96	.108
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{DF-Control}}$ × PVK_s	0.39	0.25 – 0.60	**< .001**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{DFSLM-Control}}$ × PVK_s	2.05	1.31 – 3.22	**.002**
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{VE-Control}}$ × PVK_s	1.21	0.78 – 1.87	.393
Time $_{\text{Posttest-Pretest}}$ × Group $_{\text{VESLM-Control}}$ × PVK_s	0.61	0.39 – 0.93	**.022**
Observations	15840		
Marginal R^2 / Conditional R^2	0.312 / 0.675		

Note. *ORs* = odds ratios, *CI* = confidence interval.

Table 5

Pairwise Comparisons for the Three-way Time × Group × PVK_s Interactions

Contrast	Group	PVK	*OR*	*z*	*p*
	Control	-2	1.32	1.58	.115
		0	1.51	2.68	**.007**
		2	1.73	2.28	**.023**
	DF	-2	155.22	9.54	**< .001**
		0	26.82	18.95	**< .001**
		2	4.63	4.38	**<.001**
	DFSLM	-2	3.26	2.53	**.011**
Posttest / Pretest		0	15.72	19.60	**< .001**
		2	75.80	9.54	**< .001**
	VE	-2	25.20	6.76	**< .001**
		0	42.13	24.33	**< .001**
		2	70.42	10.28	**< .001**
	VESLM	-2	104.76	9.23	**< .001**
		0	43.92	23.12	**< .001**
		2	18.41	7.64	**< .001**

Figure 7

Plot for Pre-existing Vocabulary Knowledge Moderation Effects on Meaning Recall

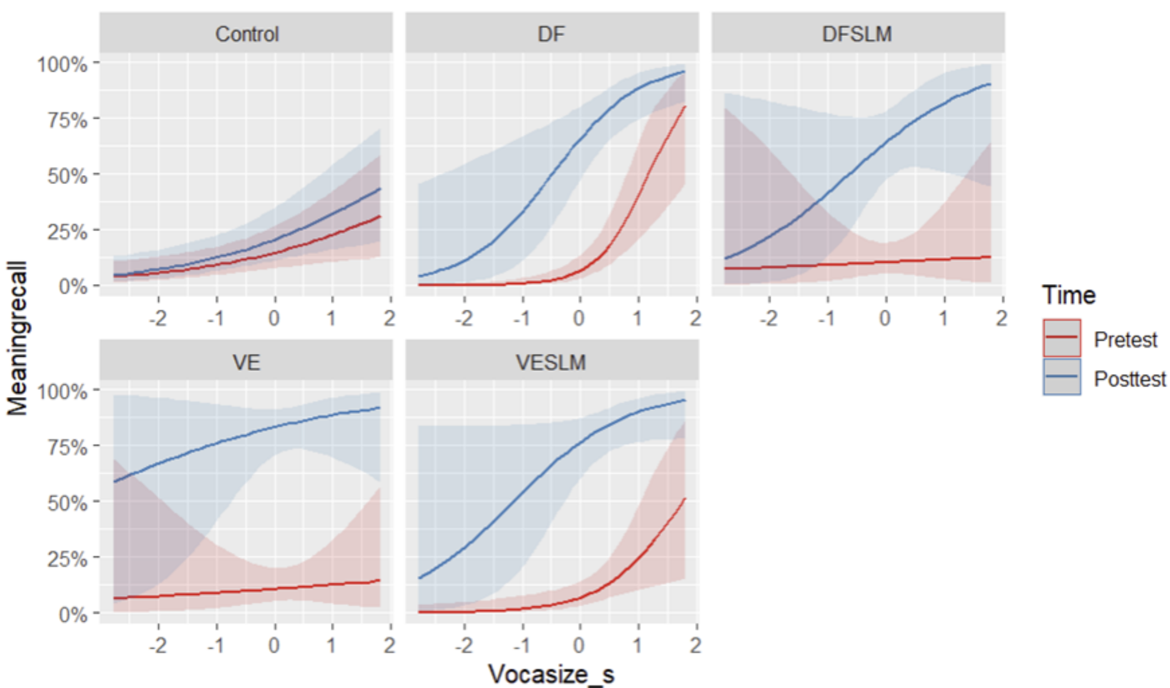
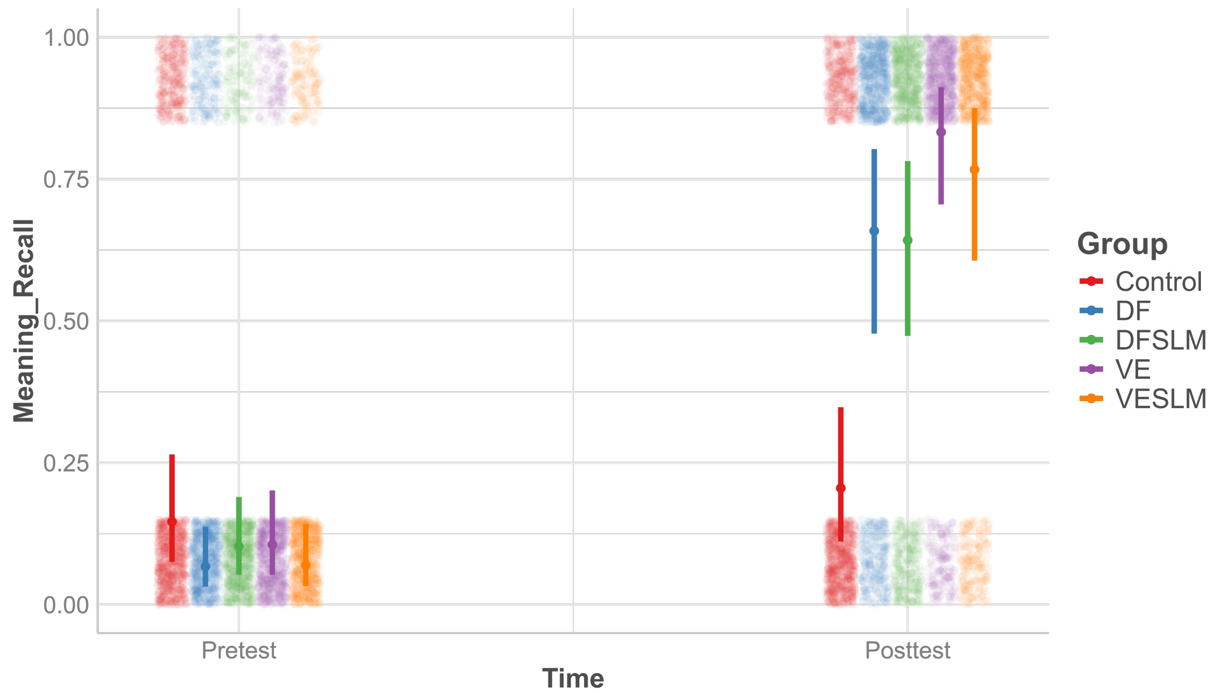


Figure 8

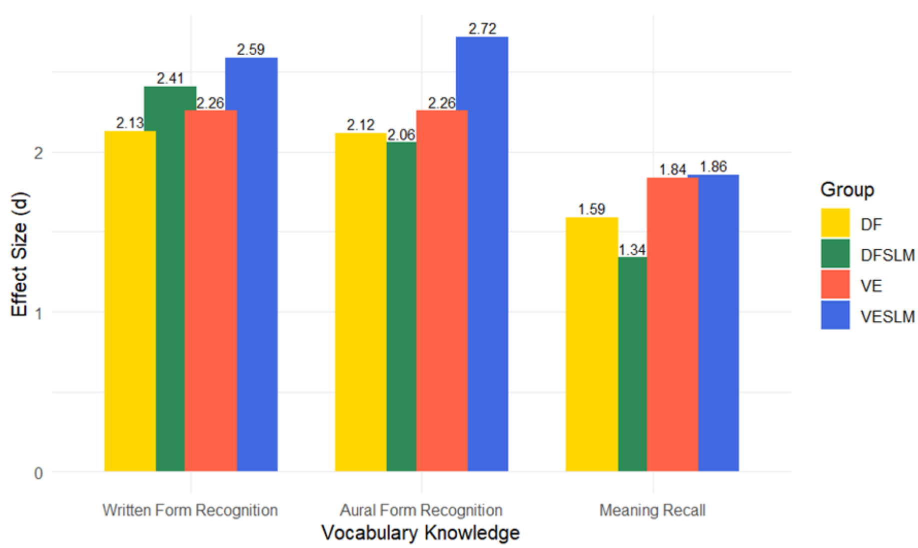
Plot for Meaning Recall Model



To conclude, all intervention groups (VESLM, DFSLM, VE, and DF) showed significantly greater gains across all three vocabulary knowledge aspects compared to the control group. To facilitate comparison with previous studies, effect sizes (odds ratios) were converted to Cohen's *d* and ranked from largest to smallest within each vocabulary knowledge aspect (Figure 9). Although DFSLM had the smallest effect on meaning recall (d = 1.34), it still met Plonsky and Oswald's (2014) criteria for a large effect (i.e., small = 0.4, medium = 0.7, large = 1.0). VESLM had the strongest overall impact on vocabulary knowledge.

Figure 9

Effect Sizes of Each Group Across Three Vocabulary Knowledge



Discussion

The Effects of Four Experimental Conditions

Based on the results, the control group made limited vocabulary gains over five weeks despite regular curriculum-based lessons, suggesting standard instruction may not support short-term vocabulary growth. Additionally, two key findings emerged: (a) both VE and DF improved vocabulary learning, with VE yielding greater overall gains; (b) the impact of SLM integration varied across vocabulary knowledge aspects. For VE, SLM enhanced written and aural form recognition but did not significantly improve meaning recall. For DF, SLM facilitated written form recognition, had no effect on aural form recognition, and slightly reduced meaning recall effectiveness.

VE significantly improved vocabulary learning, extending the findings of Mashhadi et al. (2016) and Ghoorchaei et al. (2021), who showed video instruction benefits adult learners. This study, however, focused on younger learners and reported a larger effect size (d = 1.84) compared to earlier studies (Mashhadi et al., 2016: d = 0.45; Ghoorchaei et al., 2021: d = 0.27). Unlike previous research, which assessed simpler tasks (meaning recognition), this study examined meaning recall, highlighting VE's effectiveness in deeper lexical knowledge learning. Compared to DF, VE showed superior gains across all vocabulary aspects, aligning with Paivio's dual coding theory (1986) and Mayer's (2005) cognitive theory of multimedia learning. These theories suggest that combining visual and auditory elements enhances learning by engaging multiple senses, improving attention, engagement, and retention. VE's multimodal input (video, captions) activated dual processing channels, leading to better information retention than DF's verbal-only feedback.

Additionally, VE triggered a stronger level of *evaluation*, as per the ILH (Laufer & Hulstijn, 2001), encouraging learners to contextualize words in authentic usage. However, DF relied on *search*, requiring learners to recall meanings independently, resulting in less substantial learning outcomes. This supports Yanagisawa and Webb (2021), who suggest that stronger *evaluation* within ILH may have a greater impact on vocabulary learning. Although DF was less effective than VE, it still led to significant vocabulary gains, supported by retrieval practice (Laufer et al., 2004). Actively retrieving word meanings during DF tasks strengthened memory connections. Similar benefits were reported by Yüksel et al. (2020) and Li and Hafner (2021), who found DF outperformed other methods. This study extends previous research by focusing on younger learners and demonstrating a larger effect size (d = 1.59) than those observed in adult students (Yüksel et al., 2020: d = 1.44; Li & Hafner, 2021: d = 0.46), highlighting DF's potential for early vocabulary development.

The second key finding revealed that the SLM's effect varied across vocabulary knowledge aspects. For written form recognition, both VE and DF showed greater learning gains with SLM. The improvement is likely due to the note-taking module, which encouraged word focus and strategy reflection (Chen et al., 2019). This process may have enhanced noticing of word forms (Schmidt, 1990). For aural form recognition, only VE benefited from SLM, and DF showed no difference with DFSLM. The improvement of VESLM may be attributed to note-taking, which encouraged learners to focus on word form. In VE, where both audio and visual input were provided, note-taking could also have prompted learners to reflect on the pronunciation of target words. In contrast, DF relied on meaning-focused tasks with limited non-verbal processing (Paivio, 1986). Since aural recognition depends on auditory input, SLM (goal-setting, note-taking) may not have provided enough non-verbal cues to enhance DFSLM's performance, leading to similar outcomes for DF with and without SLM on aural form recognition.

For meaning recall, VE and VESLM produced similar learning outcomes, likely because the SLM may not have offered extra support on task that is more cognitively demanding and require productive knowledge (Nation, 2022). Although SLM improved focus on word form and task engagement, it did not provide extra practice or exposure for complex vocabulary tasks like meaning recall (Swain, 1995). In contrast, DFSLM was slightly less effective than DF for meaning recall, possibly due to a conflict between DF's meaning-oriented nature and the note-taking module's focus on word form. In DF, learners selected the correct meaning from three multiple-choice options, naturally directing attention toward meaning recall. However, the note-taking module may have shifted focus to word form (e.g., spelling), diverting cognitive resources from meaning. This split attention (Paas et al., 2003) may have reduced effectiveness of DF on meaning-focused tasks, slightly lowering learning gains in meaning recall. Although previous studies (e.g., Li et al., 2021; Yüksel et al., 2020) demonstrated the benefits of VE and DF, learners externally followed pre-set instructions of VE and DF without opportunities for self-paced learning. Conversely, Chen et al. (2019) and Wang et al. (2020) examined the impact of SLM on vocabulary learning, but the learning approach (rote memorization of wordlist) lacked engaging instruction. The current study extends this line of research by integrating both VE and DF with SLM, providing clearer evidence for the effectiveness of combining technology-enhanced learning with structured self-regulated support on vocabulary learning. The SLM might support less demanding tasks (e.g., form recognition) and its effectiveness may be limited for complex tasks (e.g., meaning recall). Compared to DF, VE appears to be more suitable and effective for integration with SLM.

The Moderation Effects of Pre-Existing Vocabulary Knowledge and Self-Regulation

The lack of significant moderation by self-regulation on vocabulary knowledge across all approaches suggests that learners with varying self-regulation levels benefited similarly. This may be due to limitations of the self-report nature of how self-regulation was captured, which may not fully reveal internal factors like motivation and emotion (Zimmerman & Schunk, 2001). Although the SRLvocVLA Scale aligned with the study context, participants may have understood items differently or lacked awareness of their own states. Additionally, all interventions involved intentional, focused vocabulary learning (Webb & Nation, 2017), which may have produced strong learning effects that masked any influence of self-regulation differences. Similarly, this could explain why PVK showed no significant moderation effects on written or aural form recognition. Intentional learning directs focused attention to specific words, effectively triggering noticing (Schmidt, 1990) on written and aural forms. This strong learning effect may have overtaken any minor influence of PVK differences on form recognition gains.

For meaning recall, PVK had no significant moderation effects in VE, and VESLM showed a small negative effect ($OR = 0.61$). This may be incidental, as the large sample size increases the likelihood of detecting statistically significant but small effects. These findings suggest that VE, with or without SLM, benefits learners across different PVK levels. In contrast, DF and DFSLM showed differing trends. In DF, lower PVK learners made greater gains, whereas in DFSLM, higher PVK learners benefited more. This suggests an interaction between the difficulty of the target vocabulary, the instructional support provided, and learners' existing knowledge. SLM may have supported higher PVK learners more effectively due to its note-taking module, which allows learners to document preferred strategies (Chen et al., 2019), reinforcing retrieval practice. Their broader vocabulary knowledge likely enhanced germane cognitive load (Paas et al., 2003), helping them connect new and known words. The inclusion of lower-frequency vocabulary may have amplified this effect, offering high PVK learners more scope to apply their existing knowledge productively. In contrast, lower PVK learners may have found SLM less effective, as they had fewer lexical resources to draw on. Combined with the added demands of note-taking, the low-frequency words may have increased extraneous cognitive load for this group. This is further supported by the DF group, where without SLM, lower PVK learners performed better, likely due to reduced task complexity. Meanwhile, higher PVK learners may have found DF alone less engaging. Findings regarding PVK's moderation effects on DF and DFSLM align with El-Dakhs et al. (2022), who found that higher PVK learners used context to support vocabulary learning, while lower PVK learners struggled to do so. Zhang and Graham (2020) also showed that lower PVK learners benefited more from simpler, L1-supported instruction. These patterns support our interpretation that task complexity and instructional format interact with learners' PVK, influencing vocabulary learning outcomes.

Conclusion

This study, for the first time, redesigned and compared the impact of VE and DF on vocabulary learning within a unified digital platform, ensuring a consistent interface and controlling extraneous variables, and thereby enhancing result reliability. Compared with previous research on adult learners, this study observed greater vocabulary gains among younger learners, highlighting VE and DF's potential for early language learning. Additionally, by comparing DF (focus on *search*) and VE (focus on *evaluation*), the results confirmed that VE produced better outcomes due to its stronger emphasis on *evaluation*, reinforcing the critical role of *evaluation* in vocabulary learning (Yanagisawa & Webb, 2021).

Secondly, this study found that SLM effectively supports simpler tasks (e.g., form recognition) but has limited impact on complex tasks (e.g., meaning recall). Moreover, VE proved more suitable for SLM integration than DF. These findings contribute theoretically by emphasizing the necessity to integrate multiple frameworks for technology-enhanced vocabulary learning, as no single theory fully explains its complex, multimodal nature. While ILH (Laufer & Hulstijn, 2001) provides practical guidelines, it overlooks individual differences and lacks clear strategies for activating *need* (Yanagisawa & Webb, 2021). Therefore, the study incorporated self-regulated learning theory (Pintrich, 1995; Zimmerman & Schunk, 2001) to enhance *need*. Findings on SLM's role extend Chen et al. (2019) by demonstrating how SLM influences vocabulary learning when combined with other approaches. Finally, this study found that self-regulation did not moderate learning outcomes, but PVK influenced meaning recall in DF. SLM benefited higher PVK learners by helping them use prior knowledge, while lower PVK learners performed better with simpler tasks, such as DF without SLM. These findings contribute theoretically by integrating cognitive load theory (Paas et al., 2003) to explain how prior knowledge influences technology-enhanced vocabulary learning. The study highlights the importance of managing extraneous cognitive load, which varies by learner, and reinforcing the need to tailor instruction to cognitive capacity and prior knowledge.

This study offers practical implications for language learners, teachers, educational technology developers, and policymakers. Despite receiving regular instruction, the control group's minimal gains suggest that typical classroom practice may not sufficiently support vocabulary development. This highlights the need for more effective technology-enhanced approaches (e.g., DF, VE) for vocabulary learning. Learners with higher PVK may benefit from more comprehensive, self-regulated learning environments, whereas those with lower PVK often require more structured and straightforward input. For language teachers, adaptive digital tools can be used to deliver interactive, multimodal tasks, tailored to learners' needs. Where such tools are unavailable, self-regulated learning features can be implemented through worksheets or teacher-guided activities. Educational technology developers should prioritize the integration of personalization and support for self-regulated learning. Finally, policymakers can strengthen digital teaching competence by embedding technology use in teacher training and supporting collaboration between educators and developers to improve access and reduce inequality.

Limitation and Future Directions

Despite some limitations, the study design remains robust and offers useful insights for theory and practice. Self-regulation did not significantly affect outcomes, possibly due to limitations in the measurement tool. Although we used a questionnaire tailored to technology-based vocabulary learning, it may not have fully captured motivational and emotional factors (Zimmerman & Schunk, 2001). Future studies could incorporate interviews or observations to gain a fuller picture of learners' self-regulatory behavior. In addition, the focus on Chinese learners in an exam-driven system may limit generalizability to communication-focused contexts. Sampling across multiple schools, however, helped reduce bias and enhance the relevance of findings. Future research could examine similar interventions in diverse sociocultural settings to explore potential cross-cultural variation. Finally, the strong effects seen in the experimental groups may reflect some novelty, as learners were likely unfamiliar with VE, DF, and SLM.

Although these approaches are pedagogically grounded, future studies should examine whether such effects are sustained over time.

Acknowledgements

We are grateful to the students from the two junior secondary schools in China who took part in this study. We would also like to thank Weixin Chen for his support with participant recruitment and coordination of data collection, and Jiayang Song for her contribution to the flowchart design. We are especially thankful to Dr. Xiaobo Li for conducting the inter-rater reliability check for the qualitative data analysis.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bueno-Alastuey, M., & Nemeth, K. (2020). Quizlet and podcasts: Effects on vocabulary acquisition. *Computer Assisted Language Learning*, *35*(7), 1407–1436. https://doi.org/10.1080/09588221.2020.1802601

Chen, C., Chen, L., & Yang, S. (2019). An English vocabulary learning app with self-regulated learning mechanism to improve learning performance and motivation. *Computer Assisted Language Learning, 32*(3)*,* 237–260. https://doi.org/10.1080/09588221.2018.1485708

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Deng, Q., & Trainin, G. (2023). The effect of a self-regulated intervention on vocabulary knowledge and self-regulated learning skills for English language learners. *Reading Psychology*, *44*(7), 761–791. https://doi.org/10.1080/02702711.2023.2187908

El-Dakhs, D., Sonbul, S., & Alwazzan, R. (2022). Learning phrasal verbs in the EFL classroom: The effect of prior vocabulary knowledge and opacity. *International Review of Applied Linguistics in Language Teaching*, *60*(4), 1253–1291. https://doi.org/10.1515/iral-2020-0116

Ghoorchaei, B., Rastegar, H., Branch, G., Shomoossi, N., & Amiri, M. (2021). Podcasting and vocabulary learning in an EFL context: A mixed methods study. *Research Square*. https://doi.org/10.21203/rs.3.rs-141387/v1

Hinton, P. R., McMurray, I., & Brownlow, C. (2004). *SPSS explained* (1st ed.). Routledge.

Kızıl, A. Ş., & Savran, Z. (2018). Assessing self-regulated learning: The case of vocabulary learning through information and communication technologies. *Computer Assisted Language Learning, 31*(5–6), 599–616. https://doi.org/10.1080/09588221.2018.1428201

Kim, D., & Kim, D. J. (2012). Effect of screen size on multimedia vocabulary learning. *British Journal of Educational Technology, 43*(1), 62–70. https://doi.org/10.1111/j.1467-8535.2010.01145.x

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202–226. https://doi.org/10.1191/0265532204lt277oa

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1)*,* 1–26. https://doi.org/10.1093/applin/22.1.1

Lenth, R. (2024). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.10.6-090001) [Computer software]. https://rvlenth.github.io/emmeans/

Li, Y., & Hafner, A. C. (2021). Mobile-assisted vocabulary learning: Investigating receptive and productive vocabulary knowledge of Chinese EFL learners. *ReCALL, 34*(1), 66–80. https://doi.org/10.1017/S0958344021000161

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.004

Ministry of Education. (2017). *Compulsory education English course standard*. Beijing Normal University Press.

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford University Press.

Pintrich, P. R. (1995). Understanding self-regulated learning. *New directions for teaching and learning, 1995*(63), 3–12. https://doi.org/10.1002/tl.37219956304

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research, 45*(3), 269–286. https://doi.org/10.1080/00313830120074206

R Core Team (2024). *R: A language and environment for statistical computing* (Version 4.5.1.) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

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., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford 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

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Yanagisawa, A., & Webb, S. (2021). To what extent does the Involvement Load Hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning, 71*(2), 487–536. https://doi.org/10.1111/lang.12444

Ye, J., & Zhang, P. (2025). *Vocabulary learning through technology-enhanced learning approaches*. [Dataset]. University of Reading. https://researchdata.reading.ac.uk/1431/

Yüksel, H. G., Mercanoğlu, H. G., & Yılmaz, M. B. (2020). Digital flashcards vs. wordlists for learning technical vocabulary. *Computer Assisted Language Learning, 35*(8), 2001–2017. https://doi.org/10.1080/09588221.2020.1854312

Zhang, P., & Graham, S. (2020). Learning vocabulary through listening: The role of vocabulary knowledge and listening proficiency. *Language Learning*, *70*(4), 1017–1053. https://doi.org/10.1111/lang.12411

Zhou, K., Jin, F., Li, W., Song, Z., Huang, X., & Lin, C. H. (2024). The design of technology-enhanced vocabulary learning: A systematic review. *Education and Information Technologies, 29*, 14875–14897. https://doi.org/10.1007/s10639-023-12423-y

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Routledge.

Zou, D., Huang, Y., & Xie, H. (2019). Digital game-based vocabulary learning: Where are we and where are we going? *Computer Assisted Language Learning, 34*(5–6), 751–777. https://doi.org/10.1080/09588221.2019.1640745

Appendix A. Target Words and Word Frequency

	Word frequency results of Lextutor Vocabulary Profiler	Word frequency results of Compleat VocabProfiler
List 1		
anticipate	AWL	K-3
qualification	K2	K-2
ambition	K2	K-4
adaptation	AWL	K-2
comfort	K2	K-1
tutor	OFF	K-4
highlight	AWL	K-3
engage	OFF	K-2
involve	AWL	K-1
edition	AWL	K-3
zone	OFF	K-3
awesome	OFF	K-6
motivate	AWL	K-3
reasonable	K1	K-1
applicant	K1	K-4
List 2		
expose	AWL	K-2
firm	K2	K-2
slope	K2	K-3
insight	AWL	K-3
departure	OFF	K-4
expense	K1	K-2
dramatic	AWL	K-2
duration	AWL	K-4
surroundings	OFF	K-2
mature	AWL	K-3
depression	AWL	K-2
boom	OFF	K-2
strengthen	K1	K-2
deny	AWL	K-2
optimistic	OFF	K-4
List 3		
enrol	OFF	K-4
competent	OFF	K-3
cooperate	AWL	K-3
angle	K2	K-3
belt	K2	K-2
proceed	AWL	K-3
sincere	K2	K-4
logical	AWL	K-3
outcome	AWL	K-3
cuisine	OFF	K-6
consist	AWL	K-3

chef	OFF	K-4
elegant	OFF	K-4
exceptional	K1	K-3
minimum	AWL	K-3
List 4		
temper	K2	K-5
junk	OFF	K-5
brand	OFF	K-2
ingredient	OFF	K-3
stable	AWL	K-2
advertisement	K2	K-1
toast	OFF	K-2
calorie	OFF	K-4
category	AWL	K-3
vitamin	OFF	K-4
quantity	K1	K-3
awful	OFF	K-1
technique	AWL	K-3
urgent	OFF	K-3
chew	OFF	K-2

Appendix B. Digital Platform: Modao (https://modao.cc/)

The main functions of Modao include:

1. Adding custom images or graphics to enhance the visual design of the interface.
2. Creating multiple canvas layers with different content.
3. Adding linking buttons to connect different canvas layers and customise the display sequence.
4. Embedding external online resources such as music, videos, questionnaires, and tests.

The following example links illustrate different learning conditions using the target word *anticipate*. Since only one word was extracted as a sample, the *next one* button does not serve a functional purpose. However, in the actual experiment, the *next one* button was linked to the second target word.

Example of digital flashcards (DF): https://modao.cc/app/68BKqpiesnkyeaeaFFFIA/embed.

In the DF task, the main canvas displayed the target word and three multiple-choice options. Each option led to a different interface: one for a correct answer and others for incorrect responses with learning hints. By setting up the corresponding actions, designers can easily create interactive learning activities.

Example of video enhancement (VE): https://modao.cc/app/2uhem07snkyemNYE3QNku/embed.

The main step in designing VE was attaching a video clip to each target word. Online video clips were embedded and could be viewed when learners clicked the button.

Example of video enhancement plus self-regulated learning mechanism (VESLM):

https://modao.cc/app/k0hG5w8Rsnkyeqh5fPXtbk/embed.

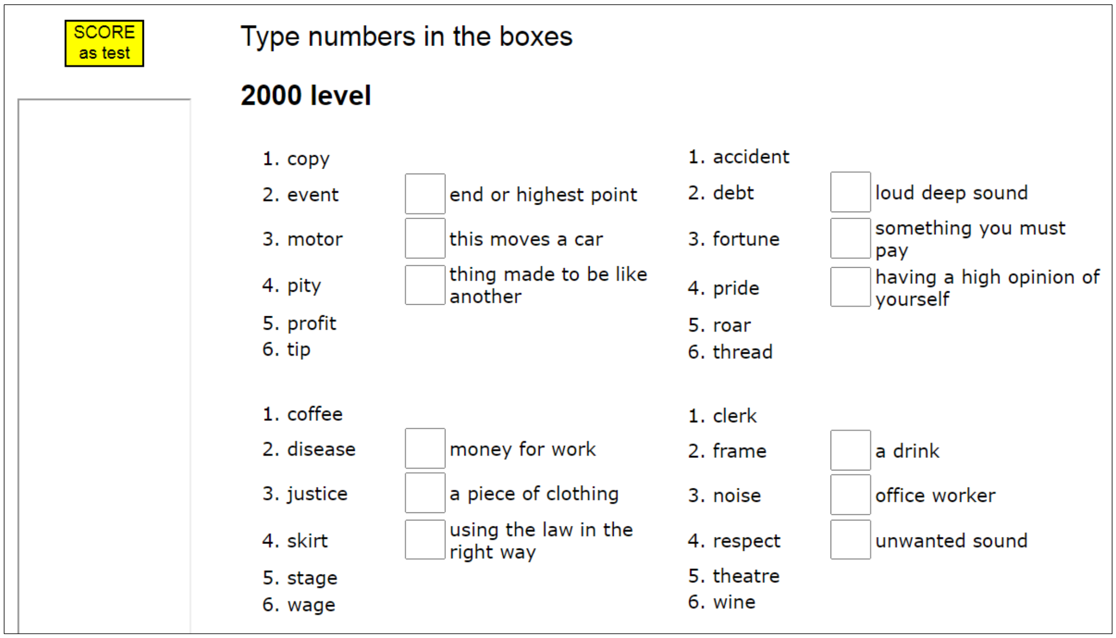
Example of digital flashcards plus self-regulated learning mechanism (DFSLM):

https://modao.cc/app/TAPDwyessnkyeiMhIEr8cn/embed.

Regarding the SLM, the goal-setting and note-taking modules are essentially two online forms, using Wenjuanxing (https://www.wjx.cn/). They were designed and attached to collect participants' expected goals and learning notes.

Appendix C. Baseline Vocabulary Levels Test

Original Version


SCORE as test

Type numbers in the boxes

2000 level

1. copy
2. event
3. motor
4. pity
5. profit
6. tip

end or highest point
this moves a car
thing made to be like another

1. accident
2. debt
3. fortune
4. pride
5. roar
6. thread

loud deep sound
something you must pay
having a high opinion of yourself

1. coffee
2. disease
3. justice
4. skirt
5. stage
6. wage

money for work
a piece of clothing
using the law in the right way

1. clerk
2. frame
3. noise
4. respect
5. theatre
6. wine

a drink
office worker
unwanted sound

Chinese Version


2000 level

* 结束或最高点： 请选择
请选择
copy
event
motor
pity
profit
tip

驱动车辆： 请选
复制： 请选择

* 咆哮： 请选择
债务： 请选择
自豪： 请选择

* 工资： 请选择
裙子： 请选择
法律制裁： 请选择

* 酒： 请选择
职员： 请选择
噪音： 请选择

Appendix D. Self-Regulation Questionnaire

Original Version

Self-regulated vocabulary learning through information and communication technologies (SRLvocICT) Scale)

	Scale Items	Not at all true of me	Somewhat not true of me	Not true of me	Somewhat true of me	True of me	Very True of me
Commitment	1. When learning vocabulary, I believe ICTs can help me achieve my goals more quickly than expected	☐	☐	☐	☐	☐	☐
	2. When learning vocabulary, I believe ICTs can help me persist until I reach the goals that I make for myself	☐	☐	☐	☐	☐	☐
	3. ICTs are important sources and tools to maintain my interest in achieving my vocabulary learning goals	☐	☐	☐	☐	☐	☐
	4. I believe ICT applications are effective in boosting willpower for learning vocabulary	☐	☐	☐	☐	☐	☐
Metacognitive	5. I know how to use ICTs to effectively monitor myself to achieve my vocabulary learning goals	☐	☐	☐	☐	☐	☐
	6. I plan tasks and relevant materials to learn vocabulary outside of school that involve the use of ICTs	☐	☐	☐	☐	☐	☐
	7. I adjust my vocabulary learning goals in response to the information resources and communication venues I have access to via ICTs	☐	☐	☐	☐	☐	☐
	8. I believe ICT tools help me monitor my progress in learning vocabulary	☐	☐	☐	☐	☐	☐
	9. I know how to adjust ICT tools according to my learning styles	☐	☐	☐	☐	☐	☐
Affective	10. During the process of learning vocabulary, I believe that ICTs can help me overcome any sense of boredom	☐	☐	☐	☐	☐	☐
	11. When feeling bored with learning vocabulary, I use ICTs to regulate my mood in order to regain the interest and enthusiasm in learning.	☐	☐	☐	☐	☐	☐
	12. When I feel stressed about vocabulary learning, I feel ICTs help to reduce this stress.	☐	☐	☐	☐	☐	☐
	13. I feel satisfied with the way I use ICTs to reduce the stress of vocabulary learning.	☐	☐	☐	☐	☐	☐
	14. I feel ICTs can make the task of vocabulary learning more attractive to me	☐	☐	☐	☐	☐	☐
	15. I feel ICTs effectively maintain my interest and enthusiasm in learning vocabulary.	☐	☐	☐	☐	☐	☐
Resource	16. When I feel I need more learning resources in vocabulary learning, I use ICTs to expand my learning resources.	☐	☐	☐	☐	☐	☐
	17. I use ICTs to create and increase opportunities to learn and use vocabulary.	☐	☐	☐	☐	☐	☐
	18. I use ICTs to seek learning resources and opportunities to help achieve my vocabulary learning goals.	☐	☐	☐	☐	☐	☐
	19. I seek engaging vocabulary learning materials and experience delivered via ICTs.	☐	☐	☐	☐	☐	☐
	20. I believe ICT tools are effective in expanding my resources for vocabulary learning.	☐	☐	☐	☐	☐	☐
Social	21. When learning vocabulary, I think ICT tools can help me create a self-paced learning environment.	☐	☐	☐	☐	☐	☐
	22. When learning vocabulary, I use ICTs to connect with native speakers of the language	☐	☐	☐	☐	☐	☐
	23. When learning vocabulary, I use ICTs to connect with peer learners all over the world	☐	☐	☐	☐	☐	☐

Revised Version

	Scale Items	0	1	2	3	4	5
Commitment	1. When learning vocabulary, I believe vocabulary learning application (VLA) can help me achieve my goals more quickly than expected. 2. When learning vocabulary, I believe VLA can help me persist until I reach the goals that I set for myself. 3. VLA is an important tool to maintain my interest in achieving my vocabulary learning goals. 4. I believe VLA is effective in boosting willpower for learning vocabulary.						
Metacognitive	5. I can use VLA to monitor myself/my progress to achieve my vocabulary learning goals. 6. I adjust my vocabulary learning goals in response to the learning resources I have access to via VLA. 7. I believe VLA can help me to monitor my progress in learning vocabulary. 8. I know how to adjust my learning methods when using VLA.						
Affective	9. During the process of learning vocabulary, I believe that VLA can help me overcome any sense of boredom. 10. When feeling bored with learning vocabulary, I believe VLA can regulate my mood in order to regain the interest in and enthusiasm for learning. 11. When I feel stressed about vocabulary learning I believe VLA can help to reduce this stress. 12. I feel satisfied with the way I use VLA to reduce the stress of vocabulary learning. 13. I feel VLA can make the task of vocabulary learning more attractive to me. 14. I feel VLA can effectively maintain my interest in and enthusiasm for learning vocabulary.						
Resource	15. I use VLA to create and increase opportunities to learn vocabulary. 16. l use VLA to seek learning resources and opportunities to help achieve my vocabulary learning goals. 17. I believe VLA are effective in expanding my resources for vocabulary learning.						

Note. 0 = Not at all true of me, 1 = Somewhat not true of me, 2 = Not true of me, 3 = Somewhat true of me, 4 = True of me, 5 = Very True of me

Appendix E. Reliability Analysis

1. Reliability analysis of the baseline vocabulary levels test

Vocabulary level	Cronbach's alpha	95% *CI*
2000 level	.90	.88 – .93
3000 level	.86	.83 – .90

2. Reliability analysis of the target vocabulary test

Pre-test

Vocabulary knowledge	**Cronbach's alpha**	**95% *CI***
Written form recognition	.92	.90 – .94
Aural form recognition	.92	.90 – .94
Meaning recall	.97	.96 – .98

Post-test

Vocabulary knowledge	**Cronbach's alpha**	**95% *CI***
Written form recognition	.99	.98 – .99
Aural form recognition	.99	.98 – .99
Meaning recall	.96	.95 – .97

3. Reliability analysis of the self-regulation questionnaire

Factors	**Cronbach's alpha**	**95% *CI***
Commitment	.92	.89 – .94
Metacognitive	.89	.85 – .92
Affective	.92	.90 – .94
Resource	.85	.80 – .89

About the Authors

Jiarun Ye recently completed his PhD at the Institute of Education, University of Reading. His research focuses on technology-enhanced vocabulary learning, self-regulated learning, and second language acquisition. He is also interested in individual differences in language learning and the role of multimodality in supporting vocabulary development and learner engagement.

E-mail: jiarun.ye@pgr.reading.ac.uk

ORCiD: https://orcid.org/0009-0008-3409-2341

Pengchong Zhang is an Associate Professor in Second Language Learning at the Institute of Education, University of Reading. His research interests include technology-enhanced language learning, school-based language learning, multimodality, vocabulary, and language comprehension. Pengchong Zhang is the corresponding author.

E-mail: anthony.zhang@reading.ac.uk

ORCiD: https://orcid.org/0000-0002-2136-4984

Rowena Kasprowicz is a UKRI Future Leaders Fellow and Associate Professor in Second Language Education in the Institute of Education at University of Reading, where she leads the Progression in Primary Languages project. Her research expertise is in classroom-based language teaching and learning with young learners.

E-mail: r.kasprowicz@reading.ac.uk

ORCiD: https://orcid.org/0000-0001-9248-6834

Catherine Tissot is Professor of Education at the Institute of Education, University of Reading. She lectures on inclusive practice and special educational needs. Her research expertise is in the field of autism and related disorders.

E-mail: c.a.tissot@reading.ac.uk

ORCiD: https://orcid.org/0000-0001-9835-0903