1. ABOUT THE DATASET

------------

Title: Vocabulary learning through technology-enhanced learning approaches

Creator(s): Jiarun Ye (https://orcid.org/0009-0008-3409-2341), Pengchong Zhang (https://orcid.org/0000-0002-2136-4984)

Organisation(s): University of Reading

Rights-holder(s): Jiarun Ye ; University of Reading

Publication Year: 2025

Description: The dataset consists of second language (English) learning data, collected from 132 Chinese learners aged 13 to 14 in two schools in China over a period of approximately 6 months. Learners were divided into four experimental groups with different technology-enhanced learning conditions (video enhancement, video enhancement plus self-regulation mechanism, digital flashcards, and digital flashcards plus self-regulation mechanism) and one control group with no intervention. The data was collected at three time points. At Time 1, quantitative data were collected through a pre-existing vocabulary knowledge test, a target vocabulary pre-test (measuring written form recognition, aural form recognition and meaning recall), and a self-regulation questionnaire. At Time 2, at the end of each learning intervention (5-week intervention, one session per week), quantitative data were collected through an immediate post-test of the target words. At time 3, quantitative data was collected one month after the last intervention session through a delayed post-test regarding the influence of repetitions under different technology-enhanced learning conditions.

Cite as: [Ye, Jiarun](https://researchdata.reading.ac.uk/view/creators/Ye=3AJiarun=3A=3A.html) and [Zhang, Pengchong](https://researchdata.reading.ac.uk/view/creators/Zhang=3APengchong=3A=3A.html) (2025): Vocabulary learning through technology-enhanced learning approaches. University of Reading. Dataset. <https://doi.org/10.17864/1947.001431>

Related publications:

Ye, Jiarun, Zhang, Pengchong, Kasprowicz, Rowena and Tissot, Catherine: The role of self-regulation and prior knowledge. Language Learning and Technology (SSCI, IF: 5.2). Under review.

Ye, Jiarun, Zhang, Pengchong, Kasprowicz, Rowena and Tissot, Catherine: Maximising vocabulary gains: The role of repetition in digital learning approaches. In progress.

Contact: jiarun.ye@pgr.reading.ac.uk

2. TERMS OF USE

------------

This dataset is licensed under a Creative Commons Attribution 4.0 International Licence: https://creativecommons.org/licenses/by/4.0/.

3. PROJECT AND FUNDING INFORMATION

------------

Title: A Comparative Study of Technology-Enhanced Vocabulary Learning Approaches: Integrating Self-Regulated Learning, Examining Repetition Effects and Role of Individual Differences

Dates: June 2025

Funding organisation: Self-funded PhD project

Grant no.:

4. CONTENTS

------------

File listing

Vocab\_technology.csv

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Description** | **Unit** | **Value labels** |
| Subject | Participant ID | Numeric | N/A |
| Group | Experimental conditions | Categorical | control vs. VE vs. VESLM vs. DF vs. DFSLM |
| Vocasize | Pre-existing vocabulary knowledge | Numeric | N/A |
| SRLlevel | Pre-existing self-regulation level | Numeric | N/A |
| Item | Target words | Categorical | N/A |
| Formrecognition | Correct or incorrect on recognising the written form | Binary | 0 = incorrect  1 = correct |
| Auralrecognition | Correct or incorrect on recognising the aural form | Binary | 0 = incorrect  1 = correct |
| Meaningrecall | Correct or incorrect meaning given | Binary | 0 = incorrect  1 = correct |
| Time | Test time points | Binary | Pretest vs. Posttest |

Vocab\_technology\_Repetitions.csv

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Description** | **Unit** | **Value labels** |
| Subject | Participant ID | Numeric | N/A |
| Group | Experimental conditions | Categorical | control vs. VE vs. VESLM vs. DF vs. DFSLM |
| Vocasize | Pre-existing vocabulary knowledge | Numeric | N/A |
| SRLlevel | Pre-existing self-regulation level | Numeric | N/A |
| Item | Target words | Categorical | N/A |
| Formrecognition | Correct or incorrect in recognising the written form | Binary | 0 = incorrect  1 = correct |
| Auralrecognition | Correct or incorrect in recognising the aural form | Binary | 0 = incorrect  1 = correct |
| Meaningrecall | Correct or incorrect meaning given | Binary | 0 = incorrect  1 = correct |
| Time | Test time points | Binary | Pretest vs. Delayed |
| Repetitions | Different numbers of repetitions (three, four, five and six) applied to the target words under each experimental condition | Numeric | 3 vs. 4 vs. 5 vs. 6 |

5. METHODS

-----------

**- Experimental procedures/protocols**

This study used a mixed-method design with 132 Chinese junior high school EFL learners from three intact classes. Participants were randomly assigned to five groups: DF (digital flashcards), DFSRL (digital flashcards + self-regulated learning mechanisms), VE (video enhancement), VESRL (video enhancement + self-regulated learning mechanisms), and a control group. Before the intervention, all participants completed a baseline vocabulary knowledge test, a self-regulation questionnaire, and a pre-test on 60 target words. Over six weeks (one 45-minute session per week), the four experimental groups studied the words under different instructional conditions. Vocabulary learning gains of the technology-enhanced learning approaches were assessed through immediate post-tests at the end of each session, targeting written form recognition, aural form recognition, and meaning recall. One month after the final session, a delayed post-test was conducted to investigate the influence of different numbers of repetitions applied to the target words under each experimental condition. The control group received no intervention but completed all measurements before the intervention (pre-test), during the final session (immediate post-test), and at the delayed post-test.

**- Environmental/experimental conditions.**

The study had four experimental conditions:

1. DF: Participants in this group completed a digital multiple-choice flashcard task for each target word, receiving immediate feedback.

2. VE: Participants in this group watched short video clips, each embedding one target word in context.

3. DFSRL: participants followed the DF formats but included SRL modules: goal-setting, note-taking, and appraisal. The goal-setting module was presented at the beginning of each session, asking learners to estimate how many words (15 words in total per session) they expected to learn. The note-taking module was integrated with each target word, providing a space for learners to record the strategies they used to remember the word. The appraisal module was administered at the end of each session as immediate post-test. It was also the target vocabulary test adopted in the current study (see Instruments used section: 3. Target vocabulary test).

4. VESRL: participants followed the VE formats but included the SRL modules mentioned above.

**- Instruments used**

1. Baseline vocabulary knowledge test

The Vocabulary Levels Test v.2 (Schmitt & Clapham, 2001) assesses vocabulary in 2000, 3000, 5000, and academic levels. As participants were junior high students with an estimated 1,600–2,000-word vocabulary (Ministry of Education, 2018), the 5000-level and academic items were excluded, and word meanings were translated into Chinese to ensure accessibility.

2. Self-regulation questionnaire

Original version: Self-Regulated Learning through Information and Communication Technologies (SRLvocICT) model proposed by Kızıl and Savran (2018)

The questionnaire measures five factors: commitment (learning goals), metacognitive (focus management), affective (emotion regulation), resource (expanding resources), and social (seeking support). To better align with this study’s objectives, peer communication and seeking support (social factor) were excluded, as participants were not allowed to interact during the experiment to prevent cross-contamination between interventions. Additionally, the wording was adapted to improve clarity for junior high school learners. The general term “information and communication technologies” was replaced with “vocabulary learning application (VLA).” After these revisions, 17 items remained.

3. Target vocabulary test

The Vocabulary Knowledge Scale by Paribakht and Wesche (1993) was adopted. Three questions per word were included: 1) “Have you seen this word before?” (written form recognition), 2) “Have you heard this word before?” (aural form recognition); 3) Providing the meaning in Chinese or English (meaning recall).

All measurements were initially developed as online assessments and then integrated into Modao (https://modao.cc/), a web-based digital platform also used to design and deliver the learning approaches (DF, DFSRL, VE, and VESRL). Learners accessed the materials via a shared web link, enabling free access across various devices. Consequently, data from the pre-existing vocabulary knowledge test, self-regulation questionnaire, and target vocabulary tests (pre-, post-, and delayed post-tests) were all collected through this platform.

**- Methods used for processing the data**

We adopted generalised liner mixed method effects models for data analysis, implemented in R (R Core Team, 2024) using the ‘lmerTest’ package (Bates et al., 2015).