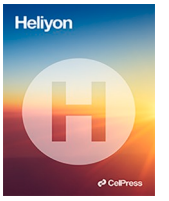
Heliyon 10 (2024) e25212

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

The facilitative effect of the keyword mnemonic on L2 vocabulary retrieval practice

Kejia Qu *, Tianzhi Liu, Yihuan Qiao, Pengcheng Wang

School of Psychology, Liaoning Normal University, Dalian, Liaoning, 116029, China

A B S T R A C T

Keyword mnemonics and retrieval practice are two learning strategies that facilitate foreign language vocabulary learning. This study examined the combination of these strategies for learning English L2 vocabulary with a limited retrieval time. We recruited 110 Chinese college students studying English as a foreign language to investigate the effects of four learning strategies on the retention of English–Chinese word pairs: restudy, retrieval practice, imposed keyword mnemonic combined with retrieval practice, and induced keyword mnemonic combined with retrieval practice. The results revealed that when retrieval practice was constrained to two times, the final performance of the retrieval practice group did not exceed that of the restudy group; however, the combined keyword-retrieval group outperformed the restudy group, regardless of whether the keyword was imposed or induced. Furthermore, there was no significant difference in memory retention performance between the induced and imposed keyword-retrieval combinations. The findings suggest that when retrieval practice is constrained to two times, the keyword-retrieval strategy combination significantly enhances English L2 vocabulary learning compared to restudy or retrieval practice alone, and both the imposed and induced keyword mnemonics can strengthen its efficiency.

1. Introduction

English is a second language (L2) that every student in China is required to master and a compulsory subject for important exams such as the Chinese College Entrance Examination. Consequently, Chinese students face immense pressure to achieve English proficiency. This raises the important question: "How can we achieve a positive effect when studying English vocabulary?", which is one of the primary objectives when learning English. Retrieval practice and keyword mnemonics are two learning strategies that may aid in vocabulary acquisition [1,2].

Retrieval practice can significantly enhance retention in the final test compared with spending the same amount of time restudying the material; this is known as the retrieval practice effect [3,4]. Several available studies demonstrate that retrieval practice, both in the laboratory and the classroom, can play a role in learners' absorption of key content [5–7]. The retrieval practice effect has been found in various materials [1,8,9] and can be applied to different age groups [8,10]. According to existing empirical studies, retrieval practice can improve foreign language vocabulary learning [11,12]. Studies on foreign words in diverse languages such as Swahili [13], Lithuanian [14], and Eskimo [15] have illustrated the benefits of retrieval practice. Retrieval practice is a practical learning approach for both university and secondary school learners seeking to develop foreign language vocabulary [16].

Keyword mnemonics is another effective memory technique to encode novel vocabulary; in this method, learners link a novel word form to its meaning with a mental image, which includes a keyword that resembles the word form [2,17]. Specifically, L2 vocabulary learning using the keyword mnemonic requires two stages: (1) the learner makes a connection between the L2 word to be learned and a word in the learner's L1 in terms of pronunciation or spelling, which serves as the keyword and (2) the learner creates a sentence that

* Corresponding author.
E-mail address: qkj0201@126.com (K. Qu).

https://doi.org/10.1016/j.heliyon.2024.e25212
Received 5 September 2023; Received in revised form 16 January 2024; Accepted 23 January 2024
Available online 24 January 2024

includes the keyword and the meaning of the foreign word in the L1 and forms an interactive image that contains both parts (i.e., an image that links the learned word pair and the keyword). This technique involves creating associations between keywords and the meanings of vocabulary items, which has been proven effective for learning foreign language vocabulary [17,18] and can help improve the associative coding between L2 words and their meanings in the L1. Thus, learners can remember the learned word pairs by refining the processing of vocabulary based on images.

1.1. Combining the keyword mnemonic and retrieval practice

Several academics have proposed integrating retrieval practice with keyword mnemonics to improve learners' vocabulary learning performance [19,20]. In practice, these two strategies are possible and easy to combine. By implementing this mnemonic technique (keywords) and subsequently practicing retrieval [21], learners can enhance their ability to remember and recall the meaning of vocabulary items. This approach taps into the power of visualization, association, and active retrieval, all of which contribute to more effective learning and retention. From a theoretical perspective, in foreign language vocabulary learning, an effective combination of keyword mnemonics and retrieval practice may enhance the encoding and retrieval of memory processes [20]. The keyword mnemonic can be viewed as encoding a mnemonic [22], and retrieval practice can be seen as a method that enhances later retrieval [23]. Keyword mnemonics can create a retrieval route to enhance memory [17].

More specifically, the mediator effectiveness hypothesis proposed by Pyc and Rawson [21] supports the effectiveness of integrating retrieval practice with keyword mnemonics, stating that testing improves memory by supporting the use of more effective keyword mediators during encoding. Participants were presented with Swahili–English translation pairs (e.g., wingu-cloud) for an initial study trial, followed by three blocks of practice trials. During the initial and restudy trials, all participants were asked to generate a keyword for each word pair. In the final test one week later, those who learned the pairs via the retrieval method were more likely to remember their keywords than those who learned them by restudying. Additionally, participants who learned the pairs through retrieval were more than twice as likely than those who learned through restudying to correctly recall the target (i.e., the English translation) when keywords were recalled during the final exam.

Miyatsu and McDaniel proposed a catalytic hypothesis that offers a new perspective on the impact of keyword-retrieval combinations [20]. Specifically, they suggest that they are superior because they combine the advantages of keyword mnemonics and retrieval practice rather than enhancing them separately. The inclusion of the keyword mnemonic during the initial study likely facilitated the benefits of retrieval practice because it offered a productive retrieval pathway that can be strengthened with further retrieval practice. They used Lithuanian–English word pairs as experimental materials and provided support for the catalytic view that the superiority of the keyword-retrieval combination is based on an interaction of the two strategies [20].

1.2. The effectiveness of combining the keyword mnemonic and retrieval practice on the retention of English L2 vocabulary

Numerous studies have examined the effects of combining retrieval practice with the keyword mnemonic on foreign language vocabulary learning and facilitating factors. However, various scholars have yet to reach a consensus on the combination effects. Some studies determined that combining retrieval practice with the keyword mnemonic did not produce better results than utilizing the two strategies alone. For example, Fritz et al. found no significant differences in the final test performance among the keyword mnemonic, retrieval practice, and a combination of the two strategies [16]. This study failed to prove the validity of combining these two strategies for the following reasons. First, the selected participants were junior high school students aged 12–13 years. The experiment used a within-subjects design in which all students had to learn the four conditions (keyword, retrieval practice, keyword combined with retrieval practice, and elaboration). Switching between strategies was difficult for this age group, which may have had fatigue effects. Second, participants were more likely to be distracted in the classroom than they would be in the laboratory.

Karpicke and Smith chose college students as their sample but did not find that combining the two strategies was effective [19]. They discovered no significant differences in the final test performance between those who learned vocabulary using a combination of the keyword mnemonic and retrieval practice and those who used repeated retrieval practice (i.e., participants learned via repeated tests until they could adequately answer all questions). This result might stem from the learning criterion in their study, which necessitated testing participants repeatedly until they recalled all target information correctly. Most notably, their results indicated that once the initial retrieval reached the criterion level, retrieval practice was sufficient, and the keyword mnemonic was no longer necessary. However, the amount of retrieval practice required for the learning criteria is often excessive, thus rendering this technique inefficient.

Similarly, the benefit of combining the two methods was not observed in Experiment 1 by Miyatsu and McDaniel [20], possibly due to the subjects' wide age range, differences in educational attainment, and an insufficiently standardized test format (the experiment was conducted on Amazon Mechanical Turk, which recruited 120 participants, $M_{\text{Age}} = 34.13$ years, range = 19–66, 51 % of whom had a bachelor's degree or higher) Furthermore, the final test in this experiment was sent via a web link, which may not guarantee an effective response. They also discovered that even when subjects were not given keywords, they spontaneously employed the mnemonic keyword. This spontaneous usage of the keyword mnemonic was a likely reason for the good performance of the control group despite having been exposed to it. The above analysis shows that certain conditions must be met for an effective combination of retrieval practice and the keyword mnemonic; namely, participants must be capable of self-learning and able to comprehend and apply both methods. Moreover, study procedures must be standardized to ensure the reliability of the results.

1.3. The effectiveness of combining the keyword mnemonic and retrieval practice with limited retrieval times

To date, the research on the retrieval practice effect in foreign language vocabulary learning has tended to use numerous test repetitions [7,9,24], such as four repetitions [11] or criterion learning, wherein participants learn by continuous restudying or taking tests until they can adequately answer all questions [25]. For example, Kang and Pashler found that in the four-time condition, the test significantly improved participants' learning performance compared to restudy, thus indicating a retrieval practice effect, whereas in the two-time condition, there was no retrieval practice effect [11]. In line with this, previous research found no retrieval practice effect when the number of retrievals in vocabulary learning was limited; that is, in two-time practice [20]. Some scholars believe that the lack of a retrieval practice effect in the presence of a small number of test repetitions is due to the randomness of connections between word pairs, with better results attained only through a large number of repeated retrieval attempts [11,20]. In this case, the keyword mnemonic incorporates the identification of a keyword and the utilization of imagery to create a strong retrieval route [17], which facilitates the retrieval practice effect.

In Experiment 2 by Miyatsu and McDaniel [20], combining retrieval practice with a keyword mnemonic yielded greater benefits than utilizing retrieval practice alone. After analyzing the results of Experiment 1, they were able to measure the use of keyword-mediated retrieval more precisely to better gauge spontaneous keyword use. The findings indicated that the two-time retrieval practice group did not significantly improve participants' final test performance compared with the restudy. When retrieval practice was combined with the keyword mnemonic, a retrieval practice effect occurred in the two-time testing condition, and the retrieval practice combined with the keyword mnemonic group performed significantly better on the final test than the restudy and retrieval practice groups.

Thus, this experiment verified the catalytic view that low dosages of retrieval practice enhance foreign language learning when combined with the keyword mnemonic [20]. Miyatsu and McDaniel used Lithuanian–English word pairs as experimental materials. Although Lithuanian differs from English in terms of pronunciation and spelling, its essential constituent letters are the same. Memory cues formed by letter pronunciation may also promote a retrieval practice effect. Chinese and English belong to different language families and have little common pronunciation or spelling. Therefore, learning English vocabulary is challenging for Chinese students. At present, whether this combined effect exists remains unclear. Focusing on the effects of retrieval practice combined with keyword mnemonics in this learning material may expand research in this area.

1.4. Methods to strengthen the efficiency of the combined strategies

Another factor to consider when the two methods are combined is the method to generate keywords, whether induced or imposed. Previous research has shown that how keywords are generated may yield different outcomes [26]. The effectiveness of the keyword method also appears to depend on the quality of the keyword images [27]; therefore, the induced keyword strategy may not be as useful for people who are not adept at forming images. Changing the way keywords are generated may enhance the strategy because they play an active role. Importantly, learning necessitates learners' active participation rather than passive information absorption. Patterns make it easier for students to connect new information with information already stored in their long-term memory [28,29]. Self-generated (induced) keywords are particularly useful because they eliminate conflicts between experimenters and subjects' encoding patterns [30]. However, a previous study involving students with academic difficulties and emotional disorders showed that learning improves more when experimenters were provided keywords than when they created them themselves [26]. Such students may struggle to find valuable keywords on their own and connect them to the interpretation of foreign words, whereas others may not face such difficulties. Therefore, it is worth investigating whether self-induced keywords are more beneficial than imposed keywords when used in conjunction with retrieval practice strategies with individuals without academic difficulties and emotional disorders.

1.5. The present study and hypotheses

The existing literature on the combination of keyword mnemonics and retrieval practice has produced inconsistent findings. Therefore, it is crucial to determine whether retrieval practice is beneficial when combined with keyword mnemonics in the context of English L2 vocabulary learning for Chinese learners. Moreover, understanding how these two strategies can be better integrated to optimize the retrieval practice effect in a constrained retrieval context is of great importance. By addressing these questions, this study aims to contribute to the existing knowledge in the field and offer insights into effective language learning strategies. Specifically, the experiment presented in this paper seeks to address the following research questions:

a. When Chinese learners learn English words in the two-time testing condition, will the strategy of combining retrieval practice with keyword mnemonics facilitate the retrieval practice effect more than retrieval practice alone?
b. Is induced keyword retrieval (i.e., participants produce keywords on their own) more effective than imposed keyword retrieval (i.e., experimenters offer keywords)?

Based on the extant literature, we built the following hypotheses:

(H1). In the two-time testing condition for English vocabulary learning, the final test performance of the retrieval practice group will not be significantly better than that of the restudy group. However, the retrieval practice combined with keyword mnemonic group will perform better in the final test than the restudy and retrieval practice groups.

(H2). The induced keyword retrieval group will perform significantly better than the imposed keyword retrieval group in the final test.

2. Materials and methods

2.1. Design and participants

The experiment was based on a factorial between-subject design. A total of 112 Liaoning Normal University undergraduate students participated in exchange for course credit for psychology courses or ¥ 15 ($M_{Age} = 20.26$; $SD = 2.15$; 85 % female). Sample size was based on Miyatsu and McDaniel's study with a power of .80 and $\alpha = 0.05$ to detect an effect of $d = 0.55$ because their Experiment 2 used stimuli, retention intervals, and initial and final tests similar to the present design [20]. Data from two participants were excluded from the analysis because they did not comply with the instructions. Students were all non-English majors and had never taken the International English Language Testing System (IELTS) test. Participants were randomly assigned to one of four conditions: SRR (study-restudy-restudy), STT (study-test-test), $K_{imposed}TT$ (imposed keyword-test-test) or $K_{induced}TT$ (induced keyword-test-test). Participants in the imposed keyword condition used the keyword mnemonic, with keywords provided by the experimenter. In the induced keyword condition, participants had to create keywords by themselves. Participants in both groups were taught the principles of keyword mnemonics and how to use them prior to the experiment. Sample sizes were as follows: SRR ($n = 26$), STT ($n = 28$), $K_{imposed}TT$ ($n = 28$), and $K_{induced}$ TT ($n = 28$). The study protocol was approved by the Ethics Committee of Liaoning Normal University.

2.2. Materials

A vocabulary syllabus was used to select concrete nouns with word lengths ranging from five to nine letters. Thirty university students who did not engage in the main experiment were asked to rate the familiarity of the words on a 7-point Likert scale (1 = "very unfamiliar" and 7 = "very familiar"). Following the evaluation, 20 pairs of English–Chinese word pairs were chosen as the experimental material, with a mean familiarity value of 1.32 ± 0.1 ranging from 1.13 to 1.5. A pre-test was administered to ascertain whether the subjects had any prior knowledge of 20 English words; they were given a list of English words and asked to write down their Chinese meanings. Participants were allowed 3 min to complete the pre-test, which was more than sufficient to complete the task. The results showed that they could not define any of the 20 English words. Additionally, before conducting further statistical analysis, we classified the 20 words based on their lengths to examine the potential impact of different word lengths. We then conducted an analysis of variance (ANOVA). The independent variable was the word length, which included four conditions: five, six, eight, and nine. The dependent variable was the scores of the corresponding words in the final test for the four groups of subjects. The results indicated no significant differences between the groups ($F(3) = 0.437$, $p = 0.73$).

2.3. Procedure

The experiment consisted of three phases: the initial study, initial retrieval, and final test (see Fig. 1). During the initial study phase,

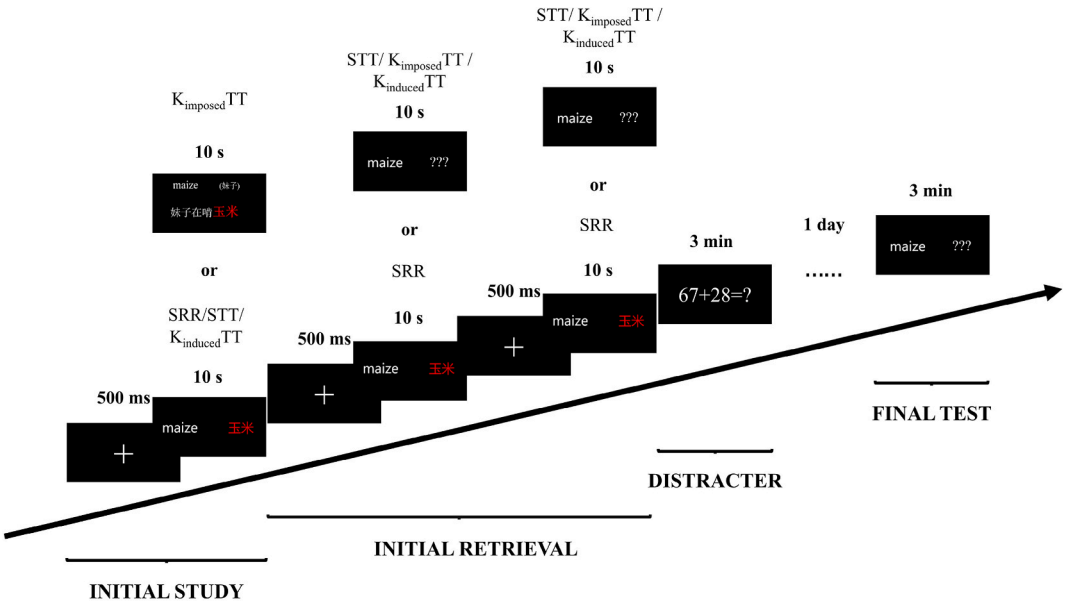


Fig. 1. The experiment procedure.

all participants were randomly shown 20 English words and their Chinese equivalents using E-prime 2.0. Participants in the STT and SRR groups were given the following instructions: "A 500 ms '+' gaze point will be provided in the center of the screen before each pair was presented, and please try to remember as many of the corresponding Chinese meaning of the English words as you can. Each pair of word will be displayed for 10 s"The $K_{imposed}$TT group and $K_{induced}$TT group differed in that, in addition to the Chinese and English word pairs, the subjects were initially given general instructions on the keyword mnemonic (i.e., what the keyword mnemonic is and how to create an image that incorporates the keyword and its Chinese meaning). Then the participants of $K_{imposed}$TT group were provided with the following instructions: "A 500 ms '+' gaze point will be provided in the center of the screen. Then English–Chinese word pair, keywords (derived from the pronunciation of the English words), and sentences linking the keywords to their Chinese pronunciation will be shown. Please mentally visualize the sentence. Each word pair will be displayed for 10 s"The $K_{induced}$TT group was given the following instructions: "A 500 ms '+' gaze point will be provided in the center of the screen. Then English–Chinese word pairs will be shown. Based on the pronunciation of the words, please try to generate keywords and create sentences connecting the keywords with the Chinese pronunciations and form a mental image of each sentence. Each word pair will be displayed for 10 s"

In the initial retrieval phase, unlike the SRR group, participants in the STT, $K_{imposed}$TT, and $K_{induced}$TT groups took two rounds of cue recall tests following the initial study, with only English words presented as cues for 10 s each. The instructions were as follows: "Please try to recall the Chinese meaning corresponding to the English word and write it down on paper." Participants in the SRR group repeatedly learned 20 English words and their Chinese equivalents twice for 10 s each. The instructions of SRR group remained the same as those in the initial learning. After the initial retrieval phase, participants engaged in a distractor task of a 3-min simple arithmetic computation. The Day 1 procedure took approximately 13 min to complete.

Participants returned to Laboratory 1D at a later time for the final cue recall test, in which they were provided with 20 English words as cues on one A4 sheet of paper from the same Chinese–English word pairs as in the initial study and asked to recollect their Chinese meanings within 3 min without feedback. The instructions remained the same as those provided to the STT group during the initial retrieval phase. Word pairs were presented randomly throughout all phases to prevent the effect of presentation order.

2.4. Results

2.4.1. Initial retrieval test score in the STT, $K_{imposed}$TT and $K_{induced}$TT conditions

To compare the initial retrieval test scores of participants in the STT, $K_{imposed}$TT, and $K_{induced}$TT treatments, a 2 (sequence: 1st vs. 2nd) × 3 (treatment: STT vs. $K_{imposed}$TT vs. $K_{induced}$TT) repeated-measures ANOVA was conducted. A manipulated between-subjects design was used for the treatment, whereas a manipulated within-subjects design was used for the sequence. There was a significant main effect for treatment ($F(2,81) = 27.462$, $p < 0.001$, $\eta_p^2 = 0.404$) but no significant main effect for sequence, $F(2,81) = 3.256$, $p = 0.075$, $\eta_p^2 = 0.039$) and no significant interaction effect between sequence and treatment ($F(2,81) = 0.756$, $p = 0.473$, $\eta_p^2 = 0.018$). Post hoc pairwise comparisons showed that the initial retrieval test scores of participants in the $K_{imposed}$TT group were significantly higher than those of the STT group ($p < 0.001$, Cohen's d = 1.17). Additionally, the initial retrieval test score of participants in the $K_{induced}$TT group was significantly higher than that of the STT group ($p < 0.001$, Cohen's d = 1.91) and higher than that of the $K_{imposed}$TT group ($p < 0.001$, Cohen's d = 0.79) (see Fig. 2). Table 1 provides descriptive statistics of the retrieval test scores in the first retrieval phase for the STT, $K_{imposed}$TT, and $K_{induced}$TT treatments, including the retrieval sequence.

2.4.2. Final test performance

To avoid the potential effects of age differences on learning outcomes, we analyzed the effects of age. The correlation analysis between age and final test scores revealed that no significant correlation (r = 0.057, $p = 0.55$). A non-parametric test was employed because the data were not normally distributed; therefore, a permutation test was used to analyze the data and obtain a more accurate result. The treatment was used as the independent variable, and the correct recall rate of Chinese vocabulary in the final test was used as the dependent variable in the permutation test, with 9999 repetitions for one-way ANOVA. The results showed that treatment had a significant main effect ($\chi^2 = 45.64$, $p < 0.001$). Post hoc pairwise comparisons revealed no significant difference between the SRR and STT groups ($p = 0.085$), as well as no significant difference between $K_{imposed}$TT and $K_{induced}$TT groups ($p = 0.179$), and participants'

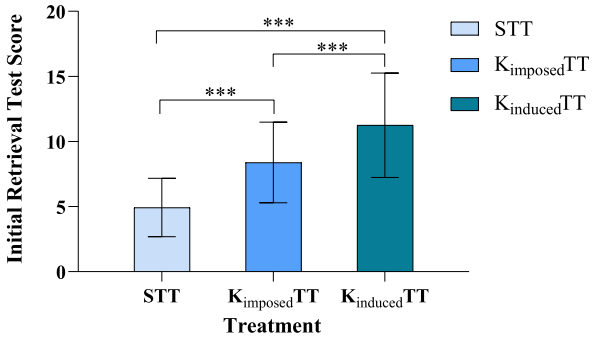


Fig. 2. Score of recalled vocabulary words in the initial retrieval test.

Table 1
Mean, standard deviation of the initial retrieval test score/the initial retrieval success rate in the STT, $K_{imposed}TT$ and $K_{induced}TT$ groups.

Treatment	First round (test score)	First round (success rate) (%)	Second round (test score)	Second round (success rate) (%)
STT	5.07 (2.12)	25.36 (10.62)	4.93 (2.24)	24.64 (11.22)
$K_{imposed}TT$	8.57 (3.68)	42.86 (18.38)	8.39 (3.10)	41.96 (15.48)
$K_{induced}TT$	11.86 (4.55)	59.30 (22.75)	11.25 (4.02)	56.25 (20.10)

Note. STT = study-test-test, $K_{imposed}TT$ = imposed keyword test, $K_{induced}TT$ = induced keyword test.

final test performance in the $K_{imposed}TT$ group was better than that of the SRR ($p = 0.014$) and STT groups ($p < 0.001$), and participants' final test performance in the $K_{induced}TT$ group was better than that of the SRR ($p = 0.002$) and STT groups ($p < 0.001$) (see Fig. 3). Descriptive statistics of the performance in the final test phase are listed in Table 2.

2.5. Discussion

In this experiment, we investigated the combined effect of keyword mnemonics and retrieval practice in acquiring English L2 vocabulary for Chinese students in a two-time testing condition and how to improve the effect. The results of the experiment supported H1 but did not support H2, implying that when retrieval practice was constrained to two times, the final performance of the retrieval practice group did not exceed that of the restudy group, but the combined keyword-retrieval groups ($K_{imposed}TT$ group and $K_{induced}TT$ group) outperformed the restudy group. Additionally, the memory retention score did not differ between the induced and imposed keyword-retrieval combinations. This study's findings suggest that combining retrieval practice with keyword mnemonics, whether induced or imposed, enhances the retrieval practice effect in English word learning compared with retrieval practice alone.

This finding contradicts that of earlier research on vocabulary learning that found retrieval practice to be more effective than restudying [13,14]. Researchers have used multiple studies or retrieval practice in previous studies to show that retrieval practice is effective in improving foreign language vocabulary learning, including four-time retrieval practice [11], multiple initial studies [9], or criterion learning [25]. However, an initial study indicates that limited retrieval practice may not consistently enhance foreign vocabulary acquisition. Kang and Pashler conducted tests on participants (either twice or four times) after a single initial exposure to Swahili–English pairs [11]. While the four-time testing condition demonstrated test-enhanced learning, the two-time testing condition did not consistently demonstrate test-enhanced learning across three experiments. These results are consistent with those of the present study. When the number of test repetitions was limited; that is, two-time practice, retrieval practice did not improve Chinese learners' English vocabulary learning performance compared with restudying.

The low initial success rate of learners in two-time retrieval practice may explain its poor performance in the subsequent final test. The different final performance of subjects in retrieval practice conditions typically depends on the level of retrieval success that students achieve in retrieval practice conditions. According to a previous meta-analysis, a stable retrieval practice effect does not exist when the retrieval success rates are below 50 % without feedback, whereas stable test effects occur when the retrieval success rates are

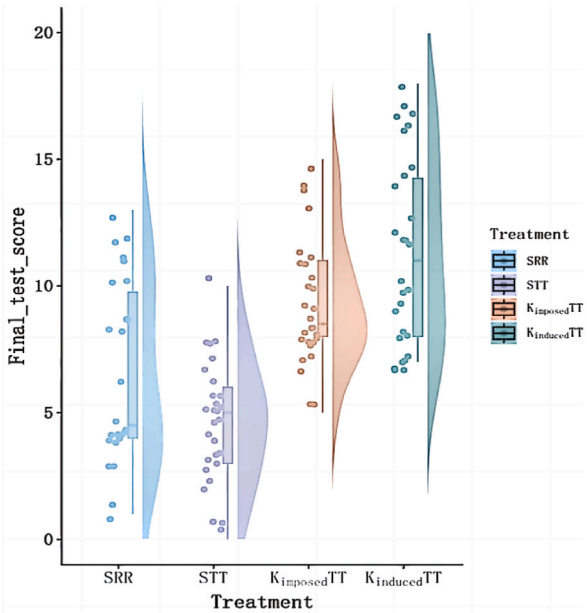


Fig. 3. Score of recalled vocabulary words in the final test. Note. SRR = study-restudy-restudy; STT = study-test-test; $K_{imposed}TT$ = imposed keyword test; $K_{induced}TT$ = induced keyword test.

Table 2
Final test score for all four treatments.

Treatment	n	*M* ± *SD*	*Md* (*P*25, *P*75)	*df*	χ^2
SRR	26	6.46 ± 3.62	4.5 (4,10)	3	45.64***
STT	28	4.64 ± 2.40	5 (3,6)		
$K_{imposed}$TT	28	9.29 ± 2.52	8.5 (8,11)		
$K_{induced}$TT	28	11.39 ± 3.75	11 (8,14.75)		

Note: Permutation test for one-way analysis of variance compares the final test performance of the four treatments. SRR = study-restudy-restudy; STT = study-test-test; $K_{imposed}$TT = imposed keyword test; $K_{induced}$TT = induced keyword test Columns: M = mean of final test score, SD = standard deviation of final test score, Md = median of final test score, P25 = first quartiles of final test score, P75 = third quartiles of final test score, χ^2 = chi-squared value of final test, df = degree of freedom of final test, *** = $p < 0.001$.

above 50 % [6]. In the present study, the initial retrieval success rate of the SRR group did not exceed 50 % (25.36 % and 24.64 % in the first and second rounds, respectively). However, when combined with keywords, the initial retrieval success rate in the $K_{imposed}$TT group increased to approximately 50 % (42.86 % and 41.96 % in the first and second rounds, respectively). The results were even higher than 50 % (59.29 % and 56.25 % in the first and second rounds, respectively) in the $K_{induced}$TT group. These results clearly demonstrate that mnemonic keywords can aid in the retrieval practice effect by increasing the initial retrieval success rate.

The findings showed that in the two-time testing condition, regardless of whether the keywords were provided by others or self-generated, when combined with keywords, the retrieval practice effect was facilitated, which is inconsistent with the findings of previous investigations [26,31]. For example, King-Sears et al. [26] discovered that keywords provided by experimenters had a better learning effect than self-generated keywords, possibly because 30 of the 37 students chosen for the study had learning difficulties and the remaining ones had emotional or behavioral disorders; thus, it was difficult for them to generate suitable keywords and independently associate them with their corresponding meanings. By contrast, the participants in our study were college students who could independently identify effective keywords and construct sentences connecting the keywords to Chinese L1 words based on their prior knowledge and experience. In addition, based on previous studies, subject generation of keywords may be more effective because it avoids conflicts between the experimenter's and subject's modes of coding [30,32]. In this approach, students are prompted to create representations of what they have learned (e.g., graphs, images, tables, or diagrams) as part of a reflection process that allows them to digest material at a deeper level, resulting in deeper memory traces and better learning outcomes [29]. However, the induced keyword strategy may not be as useful for people who are not adept at forming images. Therefore, in this study, the two methods of experimenter- and subject-generated keywords did not exhibit significant differences when combined with retrieval practice.

According to our findings, subjects in the $K_{imposed}$TT and $K_{induced}$TT groups both incorporated keyword mnemonics in their initial studies, which facilitated learners' more detailed mnemonic encoding and deeper memory traces during their initial study. Note that while the aforementioned studies all compared the differences in vocabulary learning between the two forms of keyword learning methods, our study focused on comparing the effects of these two methods combined with retrieval practice. In addition to keyword generation, retrieval practice is an important factor affecting learning outcomes in this process. Hence, despite not finding significant differences between induced and imposed keyword retrieval, the results of this study cannot conclude that there is no difference in the effectiveness of induced and imposed keywords. Based on our findings, for college students, both induced and imposed keyword retrieval were beneficial in the two-time testing condition, outperforming only retrieval practice in terms of the learning results.

These results support the catalytic view proposed by Miyatsu and McDaniel, the main concept of which is that testing alone may not effectively enhance the learning of foreign language-meaning associations owing to their arbitrary nature [20]. When there is no semantic connection between the cue (foreign vocabulary item) and the target (item meaning), one chance to study, and a limited number of practice testing rounds, the expected semantic improvement from retrieval practice may be negated. The benefits of retrieval practice were enhanced when the keyword mnemonic was used during the initial study, likely because it offered an effective retrieval pathway that could be reinforced through subsequent retrieval practice, thereby improving the likelihood of successful retrieval and facilitating the retrieval practice effect. However, the retrieval practice strategy improves the associative coding between English words and their Chinese meanings by adding keyword mnemonics, thus enabling learners to better retain what they learned in the initial study phase. Keyword mnemonics require learners to engage in a highly creative cognitive process, whereas the benefits of retrieval practice are somewhat automatic [33]. This implies that the two strategies utilize different cognitive resources and operate differently.

One potential strength of this study was the use of a controlled laboratory setting wherein participants' keyword usage patterns, whether self-generated or provided by others, can be effectively controlled to accurately reveal the effects of different types of keywords. Examining this phenomenon in natural contexts is more challenging because learners often employ various strategies rather than a specific one to achieve better learning outcomes when learning English words. In contrast, these effects can be effectively controlled in laboratory settings.

Additionally, this study utilized college students as participants. A notable characteristic of this group's learning process is its ability to autonomously identify effective keywords and construct sentences based on prior knowledge and experience. This is significant because it indicates that the participants possessed a certain level of language proficiency and cognitive ability, which are essential prerequisites for self-regulated learning. Consequently, the findings of the present study can be extrapolated to comparable cohorts.

The innovation of this study lies in investigating the combined effect of keyword mnemonics and retrieval practice in English word

learning for Chinese learners. Our findings verified that the effect of keyword-retrieval combinations applies equally to Chinese–English word learning, which belongs to different language families. Unlike most previous studies that used Swahili–English or Lithuanian–English word pairs, the experimental materials in this research were English–Chinese word pairs. Although Swahili and Lithuanian differ from English in pronunciation and spelling, their essential constituent letters are identical. However, it was difficult for the participants to memorize the words owing to the substantial variations in pronunciation and spelling between Chinese and English. Participants were unable to match the English word with its Chinese meaning due to limited learning times because the connection between the two was completely arbitrary. In the absence of semantic relations between the cue word (English word) and the target word (Chinese meaning of the word), keyword memorization enhances associative coding between words and their meanings and promotes the mastery of knowledge in the initial learning process.

This study has several limitations. First, the proportion of female participants was significantly higher because a higher proportion of female students elected to take the course. Previous research demonstrated gender differences in lexical and phonological processes. Studies using ERPs have indicated that when participants are presented with German synthetic vowels and neutral vowels during a target detection task with the instruction to respond after hearing the neutral vowels, female participants showed a stronger N100 component in the left hemisphere. Obleser. et al. and Ikezawa et al. reported similar results, suggesting that there are gender differences in lexical processing [34,35]. Therefore, the gender imbalance observed in this study likely has implications for the results. Owing to the predominance of female participants, the current overall memory performance in this study may have been higher than if there had been a balanced ratio of males to females. This finding should be considered when generalizing the results.

Second, the keyword method used in this study adopts the traditional approach of generating keywords, in which the meaning of an unfamiliar word is linked to the most phonetically similar word. However, during vocabulary acquisition, learners commonly associate the meanings of new words with terms that share similar meanings. Since this study aims to examine the beneficial impact of keyword mnemonics on L2 vocabulary retrieval practice, we utilized a widely employed method of generating keywords based on phonetic similarity. However, future research could delve deeper into this subject by examining the effects of generating keywords based on semantic similarity on foreign language vocabulary acquisition, as well as its combination with retrieval practice.

Third, as mentioned earlier, although there are benefits of conducting this research in a laboratory, it is undeniable that a laboratory setting may not fully represent a real-world classroom environment. Classroom settings involve various external stimuli and interactions that can influence students' learning experiences differently from controlled laboratory settings. For example, students' utilization of various strategies, established patterns of strategy usage, and their capacity to autonomously apply these strategies can collectively impact their learning outcomes. Therefore, this study, based on the controlled nature of the laboratory setting, limits the generalizability of its findings. Future studies should investigate the role of keyword mnemonics on the retrieval practice effect in real-world classroom teaching.

Additionally, the study participants were exclusively college students who were able to independently generate effective keywords and employ retrieval practice. However, it is important to recognize that numerous learners with diverse backgrounds or learning difficulties may encounter challenges in generating suitable keywords independently such as younger elementary school students and those with weaker self-learning abilities. Therefore, this limitation must be acknowledged when extrapolating the findings of the present study.

In summary, the results have educational implications for improving the effectiveness of English vocabulary learning and teaching. The combination of keyword mnemonics and retrieval practice facilitated the effect of learning English words in a two-time testing condition, and the combination effect was aided by both the imposed and induced keywords. These findings can enhance the effectiveness of instruction when using retrieval practice and keyword mnemonic strategies.

Ethics statement

This study was reviewed and approved by the Ethics Committee of Liaoning Normal University, with the approval number: No. LL2022024. All participants provided informed consent to participate in the study.

Data availability statement

The data can be downloaded at: https://osf.io/tr5n9/.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Kejia Qu: Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization. **Tianzhi Liu:** Writing – original draft, Visualization, Investigation, Data curation. **Yihuan Qiao:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Pengcheng Wang:** Writing – review & editing, Methodology.

Declaration of competing interest

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

Acknowledgement

This research in this article was funded by the 14th Five-years Plan of National Science of Education the Key Research Topics of the Ministry of Education (DBA230365): Research on the Effect and Mechanism of Retrieval Practice Strategy to Optimize the Worked Example Learning of Mathematical Rules for Elementary School Students.

References

[1] J. Dunlosky, K.A. Rawson, E.J. Marsh, M.J. Nathan, D.T. Willingham, Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology, Psychol. Sci. Publ. Int. 14 (1) (2013) 4–58, https://doi.org/10.1177/1529100612453266.
[2] M.E. Dikmans, G.S.E. van den Broek, J. Klatter-Folmer, Effects of repeated retrieval on keyword mediator use: shifting to direct retrieval predicts better learning outcomes, Memory 28 (7) (2020) 908–917, https://doi.org/10.1080/09658211.2020.1797094.
[3] H.L. Roediger III, J.D. Karpicke, The power of testing memory: Basic research and implications for educational practice, Perspect. Psychol. Sci. 1 (3) (2006) 181–210, https://doi.org/10.1111/j.1745-6916.2006.00012.x.
[4] S.K. Carpenter, Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval, J. Exp. Psychol. Learn. 35 (6) (2009) 1563–1569, https://doi.org/10.1037/a0017021.
[5] O.O. Adesope, D.A. Trevisan, N. Sundararajan, Rethinking the use of tests: a meta-analysis of practice testing, Rev. Educ. Res. 87 (3) (2017) 659–701, https://doi.org/10.3102/0034654316689306.
[6] C.A. Rowland, The effect of testing versus restudy on retention: a meta-analytic review of the testing effect, Psychol. Bull. 140 (6) (2014) 1432–1463, https://doi.org/10.1037/a0037559.
[7] C. Yang, L. Luo, M.A. Vadillo, R. Yu, D.R. Shanks, Testing (quizzing) boosts classroom learning: a systematic and meta-analytic review, Psychol. Bull. 147 (4) (2021) 399–435, https://doi.org/10.1037/bul0000309.
[8] A.N. Meyer, J.M. Logan, Taking the testing effect beyond the college freshman: benefits for lifelong learning, Psychol. Aging 28 (1) (2013) 142–147, https://doi.org/10.1037/a0030890.
[9] H.L. Roediger III, J.D. Karpicke, Test-enhanced learning: taking memory tests improves long-term retention, Psychol. Sci. 17 (3) (2006) 249–255, https://doi.org/10.1111/j.1467-9280.2006.01693.x.
[10] H.L. Roediger III, P.K. Agarwal, M.A. McDaniel, K.B. McDermott, Test-enhanced learning in the classroom: long-term improvements from quizzing, J. Exp. Psychol. Appl. 17 (4) (2011) 382–395, https://doi.org/10.1037/a0026252.
[11] S.H. Kang, H. Pashler, Is the benefit of retrieval practice modulated by motivation? J. Appl. Res. Mem. Cogn. 3 (3) (2014) 183–188, https://doi.org/10.1016/j.jarmac.2014.05.006.
[12] S. Candry, J. Decloedt, J. Eyckmans, Comparing the merits of word writing and retrieval practice for L2 vocabulary learning, System 89 (2020) 102206, https://doi.org/10.1016/j.system.2020.102206.
[13] J.D. Karpicke, H.L. Roediger III, The critical importance of retrieval for learning, Science 319 (5865) (2008) 966–968, https://doi.org/10.1126/science.1152408.
[14] K.E. Vaughn, K.A. Rawson, M.A. Pyc, Repeated retrieval practice and item difficulty: does criterion learning eliminate item difficulty effects? Psychon. B. Rev. 20 (2013) 1239–1245, https://doi.org/10.3758/s13423-013-0434-z.
[15] S.K. Carpenter, H. Pashler, Testing beyond words: using tests to enhance visuospatial map learning, Psychon. B. Rev. 14 (3) (2007) 474–478, https://doi.org/10.3758/BF03194092.
[16] C.O. Fritz, P.E. Morris, M. Acton, A.R. Voelkel, R. Etkind, Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning, Appl. Cognitive Psych. 21 (4) (2007) 499–526, https://doi.org/10.1002/acp.1287.
[17] R.C. Atkinson, M.R. Raugh, An application of the mnemonic keyword method to the acquisition of a Russian vocabulary, J. Exp. Psychol. Hum. Learn. Memory 1 (2) (1975) 126–133, https://doi.org/10.1037/0278-7393.1.2.126.
[18] J.A. McCabe, K.L. Osha, J.A. Roche, J.A. Susser, Psychology students' knowledge and use of mnemonics, Teach, Psychol. 40 (3) (2013) 183–192, https://doi.org/10.1177/0098628313487460.
[19] J.D. Karpicke, M.A. Smith, Separate mnemonic effects of retrieval practice and elaborative encoding, J. Mem. Lang. 67 (1) (2012) 17–29, https://doi.org/10.1016/j.jml.2012.02.004.
[20] T. Miyatsu, M.A. McDaniel, Adding the keyword mnemonic to retrieval practice: a potent combination for foreign language vocabulary learning? Mem. Cognition 47 (2019) 1328–1343, https://doi.org/10.3758/s13421-019-00936-2.
[21] M.A. Pyc, K.A. Rawson, Why testing improves memory: mediator effectiveness hypothesis, Science 330 (6002) (2010), https://doi.org/10.1126/science.1191465, 335–335.
[22] F.S. Bellezza, Mnemonic devices and memory schemas, in: M.A. McDaniel, M. Pressley (Eds.), Imagery and Related Mnemonic Processes: Theories, Individual Differences, and Applications, Springer New York, New York, NY, 1987, pp. 34–55.
[23] R.A. Bjork, E.L. Bjork, A new theory of disuse and an old theory of stimulus fluctuation, in: A. Healy, S. Kosslyn, R. Shiffrin (Eds.), From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes, ume 2, Erlbaum, Mahwah, NJ, USA, 1992, pp. 35–67. URL: https://psycnet.apa.org/record/1992-97939-014.
[24] J.D. Karpicke, H.L. Roediger III, Repeated retrieval during learning is the key to long-term retention, J. Mem. Lang. 57 (2) (2007) 151–162, https://doi.org/10.1016/j.jml.2006.09.004.
[25] J.D. Karpicke, Metacognitive control and strategy selection: deciding to practice retrieval during learning, J. Exp. Psychol. Gen. 138 (4) (2009) 469–486, https://doi.org/10.1037/a0017341.
[26] M.E. King-Sears, C.D. Mercer, P.T. Sindelar, Toward independence with keyword mnemonics: a strategy for science vocabulary instruction, Rem. Spec. Educ. 13 (5) (1992) 22–33, https://doi.org/10.1177/074193259201300505.
[27] A.A. Beaton, M.M. Gruneberg, C. Hyde, A. Shufflebottom, R.N. Sykes, Facilitation of receptive and productive foreign vocabulary learning using the keyword method: the role of image quality, Memory 13 (5) (2005) 458–471, https://doi.org/10.1080/09658210444000395.
[28] M. Wittrock, J.F. Carter, Generative processing of hierarchically organized words, Am. J. Psychol. (1975) 489–501, https://doi.org/10.2307/1421780.
[29] L. Fiorella, R.E. Mayer, Eight ways to promote generative learning, Educ. Psychol. Rev. 28 (2016) 717–741, https://doi.org/10.1007/s10648-015-9348-9.
[30] M.H. Thomas, A.Y. Wang, Learning by the keyword mnemonic: looking for long-term benefits, J. Exp. Psychol. Appl. 2 (4) (1996) 330–342, https://doi.org/10.1037/1076-898X.2.4.330.
[31] A. Campos, A. Amor, M.A. González, The importance of the keyword-generation method in keyword mnemonics, Exp. Psychol. 51 (2) (2004) 125–131, https://doi.org/10.1027/1618-3169.51.2.125.
[32] A.Y. Wang, M.H. Thomas, J.A. Ouellette, Keyword mnemonic and retention of second-language vocabulary words, J. Educ. Psychol. 84 (4) (1992) 520–528, https://doi.org/10.1037/0022-0663.84.4.520.

[33] P.E. Morris, C.O. Fritz, L. Jackson, E. Nichol, E. Roberts, Strategies for learning proper names: expanding retrieval practice, meaning and imagery, Appl. Cognitive Psych. 19 (6) (2005) 779–798, https://doi.org/10.1002/acp.1115.
[34] J. Obleser, C. Eulitz, A. Lahiri, T. Elbert, Gender differences in functional hemispheric asymmetry during processing of vowels as reflected by the human brain magnetic response, Neurosci. Lett. 314 (3) (2001) 131–134, https://doi.org/10.1016/S0304-3940(01)02298-4.
[35] S. Ikezawa, K. Nakagome, M. Mimura, J. Shinoda, K. Itoh, I. Homma, K. Kamijima, Gender differences in lateralization of mismatch negativity in dichotic listening tasks, Int. J. Psychophysiol. 68 (1) (2008) 41–50, https://doi.org/10.1016/j.ijpsycho.2008.01.006.