The Facilitating Effect of Retrieval Difficulty on the Retrieval Practice Effect for Difficult Materials: Evidence from Behavioral and fNIRS Studies

Authors: Zhang Lijuan, Jiang Yanxue, Ma Jianping, Cui Boyang, Zhang Jinkun, Zhang Jinkun

Date: 2024-10-29T15:09:53+00:00

Abstract

Research on the Retrieval Practice Effect (RPE) has revealed the critical role of retrieval in directly promoting memory retention and indirectly facilitating subsequent learning (restudy). Studies have shown that material difficulty and retrieval difficulty are important factors influencing RPE, yet previous research has not specifically distinguished their respective impacts on the direct and indirect effects of retrieval practice. Experiment 1 employed a 2 (word-pair difficulty: easy, difficult) × 2 (retrieval support: absent, present) mixed experimental design to investigate how material difficulty and retrieval difficulty affect the direct effect of retrieval practice. Experiment 2 introduced a post-retrieval restudy factor and, by combining behavioral experiments with fNIRS technology, distinguished the influences of material difficulty and retrieval difficulty on both effects of retrieval practice. The results demonstrated that, compared with the retrieval support condition, activation in the superior temporal gyrus region (associated with retrieval difficulty) during retrieval was significantly enhanced under the no-retrieval-support condition, and the prefrontal cortex region (associated with depth of processing) exhibited higher activation levels during the restudy phase, which significantly improved final memory performance. The retrieval practice effect for difficult word pairs was weaker than that for easy word pairs; however, when restudying difficult word pairs under the no-retrieval-support condition, activation levels in the prefrontal cortex region were significantly higher, effectively promoting the restudy effect following retrieval of difficult word pairs (memory performance improved significantly). These findings indicate that increasing retrieval difficulty enhances both the direct and indirect effects of retrieval practice; retrieval helps facilitate subsequent learning of difficult materials, and increasing retrieval difficulty helps strengthen the indirect effect of retrieval practice for difficult materials.

Full Text

The Facilitative Role of Retrieval Difficulty in the Retrieval Practice Effect for Difficult Materials: Evidence from Behavior and fNIRS

ZHANG Lijuan[1], JIANG Yanxue[1], MA Jianping[2], CUI Boyang[1], ZHANG Jinkun[1]
([1] School of Psychology, Fujian Normal University, Fuzhou, 350117)
([2] School of Psychology, Shandong Normal University, Jinan, 250358)

Abstract
The retrieval practice effect (RPE) demonstrates that retrieval directly enhances memory retention and indirectly facilitates subsequent learning (relearning). While material difficulty and retrieval difficulty are known to influence the RPE, previous research has not distinguished their specific impacts on the direct versus indirect effects of retrieval practice. Experiment 1 employed a 2 (word-pair difficulty: easy, difficult) × 2 (retrieval support: absent, present) mixed design to examine how material difficulty and retrieval difficulty affect the direct effect of retrieval practice. Experiment 2 introduced a post-retrieval relearning factor and combined behavioral experiments with fNIRS technology to differentiate the influences of material difficulty and retrieval difficulty on both effects. Results showed that, compared with the retrieval support condition, the no-retrieval support condition elicited significantly enhanced activation in the superior temporal gyrus (associated with retrieval difficulty) during retrieval, and the prefrontal cortex (associated with processing depth) exhibited higher activation levels during the relearning phase, which significantly improved final memory performance. The retrieval practice effect was weaker for difficult word pairs than for easy word pairs; however, when relearning difficult word pairs under no-retrieval support conditions, activation in the prefrontal cortex was significantly higher, effectively promoting the relearning effect for difficult word pairs (memory performance improved significantly). These findings indicate that increasing retrieval difficulty enhances both the direct and indirect effects of retrieval practice; retrieval facilitates subsequent learning of difficult materials, and increasing retrieval difficulty helps strengthen the indirect effect of retrieval practice for difficult materials.

Keywords: retrieval practice effect, retrieval difficulty, material difficulty, fNIRS

1 Introduction

Retrieval practice is a learning activity that strengthens memory and understanding through the recall of previously learned knowledge. Research has shown that, compared with repeated studying, retrieval practice within the same timeframe more effectively enhances learning and memory levels, a phenomenon known as the retrieval practice effect (RPE) (Roediger III & Karpicke, 2006). The RPE has been demonstrated across various learning materials, participant populations, and test formats, highlighting its stability as an efficient learning strategy across domains and age groups (McDermott, 2021). Previous research has primarily focused on the direct effect of retrieval practice—that is, its contribution to memory retention in the absence of feedback or restudying opportunities (Ma et al., 2022; Zhang & Zhang, 2020; Karpicke & Roediger, 2008). However, the broader concept of RPE also encompasses indirect effects, whereby retrieval practice enhances the accuracy of metacognitive monitoring, thereby promoting subsequent learning (Arnold & McDermott, 2013; Endres et al., 2020; Pyc & Rawson, 2012). The present study aims to explore the direct and indirect effects of retrieval practice under different conditions and their underlying neural mechanisms to deepen our understanding of the RPE and optimize efficient learning models based on retrieval practice.

The direct effect of retrieval practice reveals the critical role of retrieval in promoting memory retention. Research indicates that retrieval difficulty is an important factor influencing the magnitude of the RPE. For example, reducing retrieval cues (such as decreasing the number of letter cues during word retrieval) (Carpenter & DeLosh, 2006) or employing low-support retrieval strategies like short-answer questions (rather than multiple-choice questions) (Greving & Richter, 2022; Smith & Karpicke, 2014) increases retrieval difficulty and ultimately enhances learning and memory retention. These findings suggest that, within a certain range, greater retrieval difficulty leads to better final memory outcomes (Karpicke et al., 2014). Both the retrieval effort hypothesis and the desirable difficulty hypothesis provide theoretical explanations for this phenomenon, positing that increased retrieval difficulty prompts learners to invest more retrieval effort, which better promotes memory retention after successful retrieval (Bjork, 1975; Bjork & Bjork, 1992). Therefore, appropriately increasing retrieval difficulty has a significant positive impact on enhancing the RPE.

However, in actual learning contexts, learning content always varies in difficulty. Differences in material difficulty may also significantly influence the RPE. According to relevant theories, material difficulty itself is an important factor affecting the RPE: when learning materials are more difficult, learners must invest more retrieval effort, producing more significant memory effects (Yang et al., 2022). Carpenter (2009) found that although memory performance for easy materials (strongly associated word pairs) was better during initial retrieval, difficult materials (weakly associated word pairs) showed better performance on final tests. The elaborative retrieval account suggests that, compared with easy materials, learners activate more detailed information semantically related to target words when retrieving difficult materials (i.e., eliciting more elaborative encoding), which can serve as retrieval cues to effectively promote memory retention and retrieval success (Carpenter, 2009; Carpenter & Yeung, 2017). Furthermore, the episodic context account posits that when learning new content, learners encode both the content and its contextual information, while retrieval enables learners to update these representations and generate more effective retrieval cues (Karpicke et al., 2014). For difficult materials, retrieval involves more contextual reconstruction and updating, thereby increasing the likelihood of successful future retrieval and ultimately improving memory performance (Karpicke, 2017; Karpicke et al., 2014). Thus, material difficulty significantly impacts the RPE, with difficult materials better promoting memory retention after retrieval.

On the other hand, although numerous studies have confirmed that material difficulty influences the RPE, this effect may vary depending on retrieval difficulty (Ma et al., 2022; Smith & Karpicke, 2014). For example, Smith and Karpicke (2014) found that for easy materials, high-difficulty retrieval strategies (e.g., short-answer questions) better promoted memory retention, whereas for difficult materials, reducing retrieval difficulty (e.g., multiple-choice strategies) more effectively improved memory performance. This indicates that both material difficulty and retrieval difficulty are important factors influencing the magnitude of the RPE, and they may jointly affect the RPE. According to Sweller's (1988) cognitive load theory, total cognitive load includes intrinsic cognitive load (related to material complexity or difficulty), extraneous cognitive load (related to instructional design), and germane cognitive load (cognitive resources related to schema construction and cognitive effort). Under limited cognitive resources, effective learning requires reducing unnecessary extraneous cognitive load while ensuring necessary germane cognitive load. However, more difficult learning materials increase learners' intrinsic cognitive load and raise total cognitive load. In such cases, employing lower-difficulty retrieval strategies to appropriately reduce learners' additional cognitive load may better enhance memory effects. Although the retrieval effort hypothesis has received support (Yang et al., 2022; Kirk-Johnson et al., 2019) and suggests that difficult materials produce more significant memory effects after successful retrieval, this hypothesis may have boundary conditions. For instance, for more complex or difficult learning materials, providing retrieval support to reduce retrieval difficulty may better enhance memory effects (Karpicke et al., 2014). Therefore, the degree of direct benefit from different materials under varying retrieval difficulties and its boundary conditions remain to be further clarified.

Moreover, retrieval practice not only directly promotes learning and memory but also serves a mediating role that indirectly affects memory performance. Specifically, retrieval with feedback (i.e., restudying after retrieval) can improve learners' accuracy in assessing their own mastery, thereby promoting subsequent learning and improving memory performance (Arnold & McDermott, 2013; Pyc & Rawson, 2012; Wissman & Rawson, 2018). For example, in Arnold and McDermott's (2013) study, participants learned word pairs and then took one (1-T) or five (5-T) tests before repeated studying. After multiple test-restudy cycles, the 5-T group performed better on memory tests after restudying, and multiple prior tests (5-T group) increased the number of successfully retrieved new items on the next test. This result demonstrates that retrieval practice can effectively improve learners' subsequent learning efficiency and memory performance. However, compared with the direct effect of retrieval practice, its indirect effect has received less attention. Therefore, when exploring the effects of retrieval difficulty and material difficulty on the RPE, it is necessary to further examine and differentiate the direct and indirect effects of retrieval practice under different conditions and their underlying mechanisms. This will help us more comprehensively understand how retrieval practice affects learning and memory to further optimize learning strategies and improve learning outcomes.

A review of the literature reveals that previous research on the effect of material difficulty on the RPE has yielded inconsistent results. Some studies have shown an advantage for difficult materials in retrieval practice (Yang et al., 2022; Carpenter, 2009), while others have not observed this effect (Ma et al., 2017). Additional studies have found that final memory performance for difficult materials was worse than for easy materials, demonstrating an item difficulty effect (de Lima et al., 2020; Vaughn et al., 2013). These discrepancies may stem from study designs that did not clearly distinguish whether restudying occurred after retrieval, leading to confounding of direct and indirect effects of the RPE. For example, in Carpenter's (2009) study, participants did not receive restudying opportunities after retrieval, whereas Vaughn et al. (2013) and de Lima et al. (2020) provided restudying opportunities. Currently, it remains unclear whether the benefit of different difficulty materials in retrieval practice is influenced by post-retrieval restudying factors. Additionally, researchers have proposed that providing restudying feedback after retrieval can effectively enhance memory effects for items retrieved under low-support (high-difficulty) conditions, and offering restudying opportunities to all learners can reduce differences in initial retrieval success rates caused by varying levels of retrieval support (Kang et al., 2007). However, given that learning materials differ in difficulty, the specific impact of post-retrieval restudying on learning outcomes when employing different retrieval strategies remains unclear. Furthermore, whether the post-retrieval restudying behavioral patterns are consistent across different difficulty materials when using different retrieval strategies also awaits investigation. Therefore, further introducing the variable of post-retrieval restudying will help us more deeply understand and differentiate the specific effects of material difficulty and retrieval difficulty on the two effects of retrieval practice.

However, it should be noted that distinguishing the effects of material difficulty and retrieval difficulty on the direct and indirect effects of retrieval practice based solely on behavioral results and theoretical descriptions has certain limitations, and there remains a lack of support from cognitive neuroscience evidence. Functional magnetic resonance imaging (fMRI) studies have shown that retrieval practice is associated with activation in specific brain regions, such as the lateral temporal cortex, medial and ventrolateral prefrontal cortex (Wing et al., 2013; Ye et al., 2020; Zhuang et al., 2022), and the inferior frontal gyrus (Van den Broek et al., 2013), with activation in these regions correlating with better long-term memory performance. Moreover, research on the neural mechanisms of cognitive load has found that dorsolateral prefrontal activation is associated with intrinsic cognitive load (Whelan, 2007). Therefore, comparing activation levels in prefrontal brain regions (e.g., inferior frontal gyrus) under different retrieval difficulties and activation in dorsolateral prefrontal cortex under different material difficulties can help reveal the mechanisms underlying the effects of retrieval difficulty and material difficulty on the direct effect of retrieval practice. On the other hand, the indirect effect of retrieval practice involves enhanced encoding effort during post-retrieval restudying, and neuroimaging technology can record learners' encoding processes in real time (Nelson et al., 2013), compensating for the limitation of inferring encoding effort changes solely from subsequent retrieval performance. Additionally, previous research has confirmed the important role of prefrontal brain regions in the encoding processes that form long-term memories (Blumenfeld & Ranganath, 2007). For example, one fMRI study found that compared with learners who did not engage in retrieval practice, those in the retrieval practice group showed significantly enhanced activity in the inferior frontal gyrus and orbital frontal regions of the prefrontal cortex when restudying materials (Vestergren & Nyberg, 2014). Therefore, focusing on activation in the prefrontal cortex can help explore the cognitive neural mechanisms of post-retrieval restudying. Given fMRI's sensitivity to movement and noise, functional near-infrared spectroscopy (fNIRS) offers significant advantages for exploring cognitive activities involving movement (Kovelman et al., 2008), providing possibilities for investigating the neural mechanisms of retrieval practice in natural reading and writing recall contexts. Thus, this study intends to use fNIRS technology to further examine and differentiate the cognitive neural mechanisms underlying the two effects of retrieval practice under different conditions and provide empirical support for relevant theoretical explanations.

In summary, this study aims to explore the effects of retrieval difficulty and material difficulty on the direct and indirect effects of retrieval practice and further reveal their underlying cognitive neural mechanisms. Experiment 1 first examined the effects of retrieval difficulty and material difficulty on the direct effect of retrieval practice through a behavioral experiment. Experiment 2 further introduced the factor of post-retrieval restudying and combined behavioral experiments with fNIRS technology to examine differences in brain activation during retrieval practice and restudying under different retrieval difficulty and material difficulty conditions, thereby differentiating the effects of these two factors on the direct and indirect effects of retrieval practice. The results will provide a scientific basis for optimizing learning strategies based on retrieval practice and offer practical guidance for educational practice.

2 Experiment 1: The Direct Effect of Retrieval Practice

2.1 Purpose and Hypotheses

Experiment 1 aimed to examine the direct effects of retrieval difficulty (with or without retrieval support) and material difficulty (easy or difficult word pairs) on the RPE. We hypothesized that: (1) According to desirable difficulty theory, memory performance for difficult materials would be better than for easy materials, and memory performance in the no-retrieval support condition would be better than in the retrieval support condition; (2) According to cognitive load theory, when learning difficult materials, appropriately reducing extraneous cognitive load would better enable learners to utilize cognitive resources for information processing, so increased retrieval support should produce better learning outcomes.

2.2 Methods

2.2.1 Participants Using G*Power 3.1 software (Faul et al., 2009) to calculate sample size for a two-factor mixed design, with effect size f = 0.25, significance level $\alpha = 0.05$, and statistical power 1-$\beta = 0.8$, a total of 34 participants were required. Experiment 1 recruited 49 university students, with 25 in the no-retrieval support group and 24 in the retrieval support group. Participants had a mean age of 19.76 years (SD = 1.89). All participants were right-handed, had normal or corrected-to-normal vision, were native Chinese speakers, had not previously participated in similar psychological experiments, and received a gift upon completion.

2.2.2 Materials The learning materials consisted of paired-associate word pairs. First, 50 strongly associated and 50 weakly associated word pairs were selected from Nelson et al.'s (2004) association norms. Thirty university students who did not participate in the experiment rated the relatedness of each pair on a 7-point Likert scale (1 = lowest relatedness, 7 = highest relatedness). Thirty pairs with ratings of 5 or above were randomly selected as easy word pairs, and 30 pairs with ratings of 3 or below were selected as difficult word pairs. A paired-samples t-test on the relatedness ratings revealed that easy word pairs (M = 5.97, SD = 0.41) had significantly higher relatedness ratings than difficult word pairs (M = 2.27, SD = 0.40), $t(29) = 41.74$, $p < 0.001$, d = 9.14.

2.2.3 Design This experiment used a 2 (retrieval support: absent, present) × 2 (word-pair difficulty: easy, difficult) mixed design. Retrieval support was a between-subjects variable, and word-pair difficulty was a within-subjects variable. The dependent variables were initial retrieval recall accuracy, final test recall accuracy, and memory retention rate. Memory retention rate was calculated as: (number of items successfully retrieved at both initial retrieval and final test) ÷ (number of items successfully retrieved at initial retrieval), to examine the memory maintenance effect for initially successfully retrieved word pairs (Zhang & Zhang, 2020).

2.2.4 Procedure The experimental program was written using E-Prime 3.0 software, with word pairs presented on a computer screen. The experiment consisted of three phases: encoding, retrieval practice, and final testing, as shown in Figure 1 [Figure 1: see original paper]. There were six blocks, with three blocks each for easy and difficult word pairs. Each block contained 10 word pairs and included both encoding and retrieval phases. Blocks for easy and difficult word pairs were distributed in an interleaved fashion, and the presentation order of word pairs was balanced across blocks. Before the formal experiment, participants completed practice trials to ensure familiarity with the procedure. The specific procedure for each block was as follows:

Encoding Phase: Ten word pairs were presented sequentially. Following previous research, each pair was presented for 3 seconds (Ma et al., 2017).

Retrieval Phase: Participants retrieved the 10 learned word pairs under pseudorandom conditions. In the no-retrieval support group, "lawyer — ?" was presented at the center of the screen, and participants had 8 seconds to recall the target word based on the cue word before automatically advancing to the next pair. In the retrieval support group, "lawyer — l ?" was presented, and participants had 8 seconds to recall the target word based on both the cue and the first-letter prompt.

Final Test Phase: Two days later, participants recalled the target words for all 60 word pairs based on the cue words (e.g., "lawyer — ?" with no first-letter support for any pair). Response time was unlimited.

2.2.5 Data Analysis Behavioral data were analyzed using SPSS 29.0 statistical software. Scoring for the initial and final tests awarded 1 point for each correctly recalled word pair and 0 points for incorrect responses, with accuracy rates then calculated.

2.3 Results

2.3.1 Initial Retrieval Recall Accuracy Initial retrieval recall accuracy across retrieval support and word-pair difficulty conditions is shown in Figure 2a [Figure 2: see original paper]. A 2 (retrieval support: absent, present) × 2 (word-pair difficulty: easy, difficult) repeated-measures ANOVA revealed a significant main effect of retrieval support, $F(1, 47) = 62.97$, $p < 0.001$, ! "= 0.57, with the retrieval support group (M = 0.94, SD = 0.07) showing significantly higher initial retrieval recall accuracy than the no-retrieval support group (M = 0.70, SD = 0.20). The main effect of word-pair difficulty was also significant, $F(1, 47) = 107.32$, $p < 0.001$, ! = 0.70, with easy word pairs (M = 0.89, SD = 0.11) showing significantly higher initial retrieval recall accuracy than difficult word pairs (M = 0.74, SD = 0.22). The interaction between retrieval support and word-pair difficulty was significant, $F(1, 47) = 54.80$, $p < 0.001$, !" = 0.54.

Simple effects analysis revealed that for easy word pairs, the retrieval support group (M = 0.96, SD = 0.05) showed significantly higher initial retrieval recall accuracy than the no-retrieval support group (M = 0.83, SD = 0.11), $F(1, 47) = 29.02$, $p < 0.001$, ! "= 0.38, 95% CI = [0.08, 0.18]. For difficult word pairs, the retrieval support group (M = 0.92, SD = 0.07) also showed significantly higher initial retrieval recall accuracy than the no-retrieval support group (M = 0.58, SD = 0.18), $F(1, 47) = 73.94$, $p < 0.001$, !" = 0.61, 95% CI = [0.26, 0.43]. With increased retrieval support, difficult word pairs showed a greater improvement in initial retrieval performance than easy word pairs (0.34 vs. 0.13).

2.3.2 Final Test Recall Accuracy Final test recall accuracy across retrieval support and word-pair difficulty conditions is shown in Figure 2b. Repeated-measures ANOVA revealed a significant main effect of retrieval support, F(1, 47) = 32.53, $p < 0.001$, ! ” = 0.41, with the no-retrieval support group (M = 0.37, SD = 0.23) showing significantly higher final test recall accuracy than the retrieval support group (M = 0.22, SD = 0.20). The main effect of word-pair difficulty was significant, F(1, 47) = 239.00, $p < 0.001$, ! ” = 0.84, with easy word pairs (M = 0.47, SD = 0.18) showing significantly higher final test recall accuracy than difficult word pairs (M = 0.12, SD = 0.09). The interaction between retrieval support and word-pair difficulty was not significant, F(1, 47) = 3.70, p = 0.06.

2.3.3 Memory Retention Rate Memory retention rates across retrieval support and word-pair difficulty conditions are presented in Table 1 . Repeated-measures ANOVA revealed a significant main effect of retrieval support, F(1, 47) = 68.05, $p < 0.001$, ! ” = 0.59, with the no-retrieval support group (M = 0.49, SD = 0.24) showing significantly higher memory retention rates than the retrieval support group (M = 0.22, SD = 0.20). The main effect of word-pair difficulty was significant, F(1, 47) = 181.13, $p < 0.001$, ! ” = 0.79, with easy word pairs (M = 0.53, SD = 0.22) showing significantly higher memory retention rates than difficult word pairs (M = 0.19, SD = 0.17). The interaction was not significant, F(1, 47) = 1.28, p = 0.26.

Table 1 Memory Retention Rates by Retrieval Support and Word-Pair Difficulty

	No Retrieval Support	Retrieval Support
Easy Word Pairs	0.67 (0.16)	0.30 (0.15)
Difficult Word Pairs	0.38 (0.16)	0.07 (0.06)

Note: Standard deviations are shown in parentheses.

2.4 Discussion

Experiment 1 results showed that during the initial retrieval phase, the retrieval support group demonstrated significantly higher memory performance than the no-retrieval support group. However, this pattern reversed on the final test two days later, with the no-retrieval support group showing significantly better memory performance than the retrieval support group. This finding is consistent with Carpenter and DeLosh (2006) and indicates that although retrieval support improves initial retrieval performance, it may accelerate forgetting and is not conducive to long-term memory retention, supporting the retrieval effort hypothesis and desirable difficulty theory.

However, compared with easy word pairs, difficult word pairs showed poorer memory performance on both initial retrieval and final testing, demonstrating an item difficulty effect. This result is inconsistent with previous findings showing an advantage for difficult items in the RPE (Yang et al., 2022; Carpenter, 2009). According to RPE theories, learners invest more retrieval effort when facing more difficult items, thereby promoting memory retention. However, Experiment 1 suggests that the advantage of difficult materials in the RPE is not significant. According to cognitive load theory (Sweller, 1988), learners experience greater intrinsic cognitive load when processing difficult word pairs, which may lead to poorer encoding or retrieval failure, ultimately affecting memory performance. Additionally, Vaughn et al. (2013) found that the item difficulty effect persisted even when the number of successful retrievals was equal for easy and difficult materials. They speculated that this might be because difficult word pairs struggle to form effective associative memories, making target words difficult to retrieve successfully. Therefore, for difficult word pairs, merely improving initial retrieval success rates may not be sufficient to effectively promote memory retention; investing additional learning time after retrieval practice to strengthen associative memory may be more crucial. Research shows that post-retrieval restudying helps improve learning outcomes, especially for items that failed initial retrieval (Arnold & McDermott, 2013; de Lima et al., 2020). Combined with Experiment 1 results, it appears that simply relying on retrieval support to improve initial retrieval success rates for difficult word pairs may not be the key to enhancing overall learning outcomes; instead, strengthening subsequent encoding effort may be more important.

Overall, Experiment 1's findings and explanations regarding the effects of retrieval support and word-pair difficulty on the RPE remain at the behavioral data level. To further verify and explore: (1) whether learners invest more retrieval effort when using no-retrieval support strategies; (2) whether difficult materials create greater intrinsic cognitive load that affects final memory outcomes; and (3) whether post-retrieval restudying better improves learning outcomes for difficult materials, Experiment 2 employed fNIRS technology for verification and investigation, providing a neural basis for these results and explanations. Additionally, Experiment 2 introduced the factor of post-retrieval restudying and combined behavioral experiments with fNIRS technology to further examine the effects of retrieval difficulty and material difficulty on both direct and indirect effects of retrieval practice and their neural mechanisms.

3 Experiment 2: Direct and Indirect Effects of Retrieval Practice

3.1 Purpose and Hypotheses

Based on Experiment 1 results and relevant theories, we propose the following hypotheses for behavioral results: (1) The RPE results for the no-restudy group (direct effect) would be consistent with Experiment 1; (2) If items that failed retrieval benefit more from post-retrieval restudying, then under restudy conditions, difficult word pairs would show greater benefit than under no-restudy conditions, particularly for difficult word pairs under no-retrieval support conditions.

For fNIRS results, we hypothesized: (1) Compared with retrieval support conditions, retrieval under no-retrieval support conditions would elicit significantly higher activation in prefrontal cortex regions closely related to retrieval activity (e.g., inferior frontal gyrus); (2) Compared with easy word pairs, retrieval of difficult word pairs would significantly enhance activation in dorsolateral prefrontal cortex regions associated with cognitive load; (3) During the post-retrieval restudying phase, activation in prefrontal cortex regions (e.g., inferior frontal gyrus related to re-encoding) would be stronger under no-retrieval support conditions than under retrieval support conditions, and activation would be stronger when restudying difficult word pairs than when restudying easy word pairs.

3.2 Methods

3.2.1 Participants Using G*Power 3.1 software (Faul et al., 2009) to calculate sample size for a three-factor mixed design, with effect size f = 0.25, significance level $\alpha = 0.05$, and statistical power 1-$\beta = 0.8$, a total of 48 participants were required. Experiment 2 recruited 68 university students who were randomly assigned to one of four groups: the post-retrieval restudy group (n = 33; 16 in retrieval support condition, 17 in no-retrieval support condition) and the no-restudy group (n = 35; 17 in no-retrieval support condition, 18 in retrieval support condition). fNIRS data were not successfully collected for one participant in the no-retrieval support condition of the restudy group and one participant in the retrieval support condition of the no-restudy group (both withdrew mid-experiment). Participants had a mean age of 21.74 years (SD = 2.47). Other participant characteristics were the same as in Experiment 1.

3.2.2 Materials Same as Experiment 1.

3.2.3 Design This experiment used a 2 (retrieval support: absent, present) × 2 (word-pair difficulty: easy, difficult) × 2 (post-retrieval restudy: absent, present) three-factor mixed design. Retrieval support and post-retrieval restudy were between-subjects variables, and word-pair difficulty was a within-subjects variable.

3.2.4 Procedure The procedure for the no-restudy group was the same as in Experiment 1. However, in the fNIRS experiment, rest periods were required between tasks to ensure that blood oxygen levels returned to baseline. Therefore, following Wan et al. (2018), a 30-second rest period was set before and after each block, and a 10-second rest period was set between different tasks within the same block. The experimental procedure for the no-restudy condition is shown in Figure 3 [Figure 3: see original paper] (solid-line boxes).

The difference between the restudy and no-restudy conditions was that participants in the restudy group received restudying opportunities after completing the encoding and retrieval tasks in each block. After a 10-second rest, they restudied the materials, then rested for 30 seconds before entering the retrieval practice task for the next block. Other aspects were the same as in the no-restudy group. The experimental procedure for the restudy condition is shown in Figure 3 (dashed-line boxes).

3.3 fNIRS Data Collection

Brain activity data were recorded using a LIGHTNIRS portable functional near-infrared spectroscopy system (fNIRS Devices LLC, Potomac, MD) produced by Shimadzu Corporation. Based on relevant literature, this study focused on the prefrontal cortex. To cover relevant brain regions, a 2 × 8 channel arrangement was used, totaling 22 channels, with a 3-cm distance between sources and detectors (see Figure 4 [Figure 4: see original paper]). Probe placement was based on the international 10-20 EEG system (Okamoto et al., 2004). A 3D digitizer (FASTRAK, Polhemus, Colchester, VT, USA) was used for positioning, and Matlab 2014a and NIRS_{SPM} (Near infrared Spectroscopy-Statistical Parametric Mapping) software were used to register each location to MNI (Montreal Neurological Institute) space coordinates using probabilistic registration to determine the correspondence between each channel location and Brodmann areas (BA). The Brodmann areas corresponding to each channel are shown in Table 2 .

Figure 4 fNIRS Channel Layout

Table 2 Brodmann Areas Corresponding to fNIRS Channels

Channel	Brodmann Area	Brain Region
1	47	Left inferior frontal gyrus
2	45	Left inferior frontal gyrus, pars triangularis
3	47	Left inferior frontal gyrus
4	10	Left frontal pole
5	10	Left frontal pole
6	10	Frontal pole
7	11	Right orbital frontal cortex
8	10	Right frontal pole
9	47	Right inferior frontal gyrus
10	47	Right inferior frontal gyrus
11	6	Left premotor and supplementary motor cortex
12	22	Left superior temporal gyrus
13	45	Left inferior frontal gyrus, pars triangularis
14	46	Left dorsolateral prefrontal cortex
15	10	Left frontal pole
16	10	Left frontal pole

Channel	Brodmann Area	Brain Region
17	10	Right frontal pole
18	10	Right frontal pole
19	46	Right dorsolateral prefrontal cortex
20	45	Right inferior frontal gyrus, pars triangularis
21	22	Right superior temporal gyrus
22	9	Left dorsolateral prefrontal cortex

Note: Only brain regions with coverage probability ≥ 30% are reported.

3.4 Data Processing

Behavioral data were processed as in Experiment 1. fNIRS data were preprocessed using NIRS_{SPM} software in Matlab 2014a. (1) A hemodynamic response function (HRF) was used to remove high-frequency noise while preserving signals consistent with hemodynamic response characteristics. (2) Wavelet analysis (Wavelet Minimum Description Length, Wavelet-MDL) was used for filtering and detrending based on the distribution of experimental stimulation signals to correct for effects of respiration, participant movement, physiological noise, and equipment noise on fNIRS data. These methods have been validated in previous research (Yang et al., 2019; Brigadoi et al., 2014). Additionally, research has found that oxygenated hemoglobin (HbO) is relatively sensitive to task stimulation changes (Hoshi et al., 2001; Pinti et al., 2020). Therefore, HbO concentration change values (Δ[HbO]) were used as the analysis metric for fNIRS results.

After preprocessing, β values related to task conditions were calculated using the General Linear Model (GLM) as a measure of brain region activation. Reference waves for HbO changes were set for each channel to evaluate the data, and β values for the initial encoding phase, retrieval phase, and post-retrieval restudying phase were calculated separately. Statistical analyses were conducted using SPSS 29.0 software, with repeated-measures ANOVA, independent-samples t-tests, and paired-samples t-tests performed on β values for HbO changes according to experimental purposes. The false discovery rate (FDR) was used to correct p-values across channels to further reduce false positive rates, with $p < 0.05$ considered significant after correction (Noble, 2009). The xjview toolbox (https://www.alivelearn.net/xjview) and BrainNet Viewer software (http://www.nitrc.org/projects/bnv/) (Xia et al., 2013) were used to visualize brain imaging analysis results.

3.5 Behavioral Results

To more intuitively examine the effect of retrieval practice on final memory outcomes, only final test recall accuracy was analyzed as the dependent variable.

3.5.1 Behavioral Results for Direct Effects of Retrieval Support and Word-Pair Difficulty According to experimental purposes, we first analyzed the effects of retrieval support and word-pair difficulty on final test recall accuracy under no-restudy conditions. Consistent with Experiment 1 results, a 2 (retrieval support: absent, present) × 2 (word-pair difficulty: easy, difficult) repeated-measures ANOVA revealed a significant main effect of retrieval support, $F(1, 33) = 11.68$, $p = 0.002$, ! ” = 0.26, with the no-retrieval support group ($M = 0.36$, $SD = 0.25$) showing significantly higher final test recall accuracy than the retrieval support group ($M = 0.21$, $SD = 0.19$). The main effect of word-pair difficulty was significant, $F(1, 33) = 174.54$, $p < 0.001$, ! ” = 0.84, with easy word pairs ($M = 0.44$, $SD = 0.21$) showing significantly higher final test recall accuracy than difficult word pairs ($M = 0.13$, $SD = 0.11$). The interaction between retrieval support and word-pair difficulty was not significant, $F(1, 33) = 3.32$, $p = 0.08$.

3.5.2 Behavioral Results for Indirect Effects of Retrieval Support and Word-Pair Difficulty According to experimental purposes, we further analyzed the effects of retrieval support and word-pair difficulty on the indirect effect of retrieval practice under restudy conditions. A 2 (retrieval support: absent, present) × 2 (word-pair difficulty: easy, difficult) repeated-measures ANOVA revealed a significant main effect of retrieval support, $F(1, 31) = 18.71$, $p < 0.001$, ! ” = 0.38, with the no-retrieval support group ($M = 0.47$, $SD = 0.25$) showing significantly higher final test recall accuracy than the retrieval support group ($M = 0.22$, $SD = 0.21$). The main effect of word-pair difficulty was significant, $F(1, 31) = 85.30$, $p < 0.001$, ! ” = 0.73, with easy word pairs ($M = 0.49$, $SD = 0.24$) showing significantly higher final test recall accuracy than difficult word pairs ($M = 0.21$, $SD = 0.20$). The interaction between retrieval support and word-pair difficulty was not significant, $F(1, 31) = 0.31$, $p = 0.58$.

Additionally, to further examine whether post-retrieval restudying better improves memory performance for difficult word pairs, we separately analyzed the effects of word-pair difficulty and post-retrieval restudying on memory performance under no-retrieval support and retrieval support conditions.

No-Retrieval Support Condition: A 2 (post-retrieval restudying: absent, present) × 2 (word-pair difficulty: easy, difficult) repeated-measures ANOVA on final test recall accuracy (see Figure 5a [Figure 5: see original paper]) revealed no significant main effect of post-retrieval restudying, $F(1, 32) = 3.16$, $p = 0.09$, with no significant difference between the restudy group ($M = 0.47$, $SD = 0.25$) and the no-restudy group ($M = 0.36$, $SD = 0.25$). The main effect of word-pair difficulty was significant, $F(1, 32) = 162.36$, $p < 0.001$, ! ” = 0.84, with easy word pairs ($M = 0.58$, $SD = 0.20$) showing significantly higher final test recall accuracy than difficult word pairs ($M = 0.25$, $SD = 0.191$). The interaction between post-retrieval restudying and word-pair difficulty was not significant, $F(1, 32) = 1.01$, $p = 0.31$.

According to experimental purposes and following previous research (Zhang et al., 2020), we further conducted planned comparisons (Shu & Zhang, 2008) using independent-samples t-tests to compare final test recall accuracy for different difficulty word pairs between restudy and no-restudy conditions under no-retrieval support. Results showed that for easy word pairs, there was no significant difference in final test recall accuracy between the restudy group (M = 0.62, SD = 0.19) and the no-restudy group (M = 0.54, SD = 0.21), t(1, 32) = 1.16, p = 0.25. For difficult word pairs, the restudy group (M = 0.32, SD = 0.22) showed significantly higher final test recall accuracy than the no-restudy group (M = 0.18, SD = 0.13), t(1, 32) = 2.15, p = 0.04, d = 0.77, 95% CI = [-0.01, 0.26].

Retrieval Support Condition: A 2 (word-pair difficulty: easy, difficult) × 2 (post-retrieval restudying: absent, present) repeated-measures ANOVA on final test recall accuracy (see Figure 5b) revealed a significant main effect of word-pair difficulty, F(1, 32) = 87.55, p < 0.001, ! ” = 0.73, with easy word pairs (M = 0.35, SD = 0.19) showing significantly higher final test recall accuracy than difficult word pairs (M = 0.08, SD = 0.06). The main effect of post-retrieval restudying was not significant, F(1, 32) = 0.16, p = 0.69, with no significant difference between the restudy group (M = 0.22, SD = 0.21) and the no-restudy group (M = 0.21, SD = 0.19). The interaction between post-retrieval restudying and word-pair difficulty was not significant, F(1, 32) = 0.001, p = 0.97.

Planned comparisons using independent-samples t-tests were conducted to compare final test recall accuracy for different difficulty word pairs between restudy and no-restudy conditions under retrieval support. Results showed that for easy word pairs, there was no significant difference between the restudy group (M = 0.36, SD = 0.21) and the no-restudy group (M = 0.34, SD = 0.17), t(1, 32) = 0.23, p = 0.82. For difficult word pairs, there was no significant difference between the restudy group (M = 0.09, SD = 0.08) and the no-restudy group (M = 0.07, SD = 0.05), t(1, 32) = 0.78, p = 0.44.

Figure 5 Final test recall accuracy for different word-pair difficulty and post-retrieval restudying conditions under no-retrieval support (Figure 5a) and retrieval support (Figure 5b)

3.5.3 Effects of Retrieval Support, Word-Pair Difficulty, and Post-Retrieval Restudying on the RPE

Additionally, we conducted a 2 (retrieval support: absent, present) × 2 (word-pair difficulty: easy, difficult) × 2 (post-retrieval restudying: absent, present) three-factor repeated-measures ANOVA. Results revealed a significant main effect of retrieval support, F(1, 64) = 30.74, p < 0.001, ! ” = 0.32, with the no-retrieval support group (M = 0.42, SD = 0.25) showing significantly higher final test recall accuracy than the retrieval support group (M = 0.22, SD = 0.20). The main effect of word-pair difficulty was significant, F(1, 64) = 239.59, p < 0.001, ! ” = 0.80, with easy word pairs (M = 0.46, SD = 0.23) showing significantly higher final test recall accuracy than difficult word pairs (M = 0.17, SD = 0.16). The main effect of post-retrieval restudying was not significant, F(1, 64) = 2.86, p = 0.10, with no significant difference between the restudy group (M = 0.35, SD = 0.26) and the no-restudy group (M = 0.28, SD = 0.23). The three-way interaction between retrieval support, word-pair difficulty, and post-retrieval restudying was not significant, $F(1, 64) = 0.45$, $p = 0.51$, and none of the two-way interactions were significant ($Fs(1, 64) < 2.44$, $ps > 0.12$).

3.6 fNIRS Results

3.6.1 Brain Activation Results for Direct Effects of Retrieval Support and Word-Pair Difficulty A 2 (retrieval support: absent, present) × 2 (word-pair difficulty: easy, difficult) repeated-measures ANOVA was conducted on β values for HbO concentration changes during the retrieval phase in the no-restudy group. Results revealed a significant main effect of retrieval support in channel 15, with the no-retrieval support group showing stronger cortical activity than the retrieval support group. The main effect of word-pair difficulty was significant in channels 2, 9, 14, 16, and 17 ($p < 0.05$ before correction), with difficult word pairs eliciting stronger cortical activity than easy word pairs. The interaction between retrieval support and word-pair difficulty was not significant in any channel ($Fs(1, 32) < 1.55$, $ps > 0.22$). Detailed information about significant channels is presented in Table 3 , and brain activation maps are shown in Figure 6 [Figure 6: see original paper].

Table 3 Significant Channels for Main Effects of Retrieval Support and Word-Pair Difficulty During Retrieval Phase

Effect	Channel	Brodmann Area	Brain Region
Retrieval Support	15	22	Right superior temporal gyrus
Word-Pair Difficulty	2	47	Left inferior frontal gyrus
	9	10	Left frontal pole
	14	45	Left inferior frontal gyrus, pars triangularis
	16	46	Left dorsolateral prefrontal cortex
	17	46	Right dorsolateral prefrontal cortex

Figure 6 Brain activation maps for main effects of retrieval support and word-pair difficulty during retrieval phase (Note: Color figures available in electronic version)

3.6.2 Brain Activation Results for Indirect Effects of Retrieval Support and Word-Pair Difficulty According to experimental purposes, we first compared brain activation intensity during the post-retrieval restudying phase across different retrieval support and word-pair difficulty conditions to examine the cognitive mechanisms underlying the indirect effect of retrieval practice. We then further compared brain activation during initial learning and post-retrieval restudying phases for different word-pair difficulty conditions under retrieval support and no-retrieval support conditions to verify the behavioral results for the indirect effect of retrieval practice. Specifically, under no-retrieval support conditions, providing restudying opportunities should help improve memory performance for difficult word pairs.

Comparison of Brain Activation Intensity During Post-Retrieval Restudying Across Retrieval Support and Word-Pair Difficulty Conditions

Given that different difficulty materials retrieved using different strategies show differences in initial retrieval success rates, which may lead to inconsistent encoding effort during restudying, we adopted Darnai et al.'s (2019) method to use the difference in brain activity between two related tasks as the activation intensity value for one task. Specifically, we used the difference between β values for HbO concentration changes during post-retrieval restudying and initial learning phases as the brain activation intensity value for the post-retrieval restudying phase.

A 2 (retrieval support: absent, present) $\times$ 2 (word-pair difficulty: easy, difficult) repeated-measures ANOVA on brain activation intensity values during the post-retrieval restudying phase in the restudy group revealed significant main effects of retrieval support in channels 12 and 18, with the no-retrieval support group showing significantly higher brain activation intensity during restudying than the retrieval support group. The main effect of word-pair difficulty was significant in channels 1, 7, 9, 16, 17, 20, and 21, with difficult word pairs showing significantly higher brain activation intensity during restudying than easy word pairs. The interaction between retrieval support and word-pair difficulty was not significant in any channel ($Fs(1, 30) < 2.53$, $ps > 0.12$). Detailed information about significant channels is presented in Table 4 , and brain activation maps are shown in Figure 7 [Figure 7: see original paper].

Table 4 Significant Statistical Results for Main Effects of Retrieval Support and Word-Pair Difficulty During Post-Retrieval Restudying Phase

Effect	Channel	Brodmann Area	Brain Region	F (1, 31)
Retrieval Support	12	10	Right frontal pole	6.45
	18	10	Left frontal pole	5.12
Word-Pair Difficulty	1	47	Left inferior frontal gyrus	7.23
	7	45	Left inferior frontal gyrus, pars triangularis	8.91
	9	47	Right inferior frontal gyrus	6.78
	16	9	Left dorsolateral prefrontal cortex	9.34
	17	9	Left dorsolateral prefrontal cortex	7.56
	20	10	Right frontal pole	5.89
	21	46	Right dorsolateral prefrontal cortex	8.12

Figure 7 Brain activation maps for main effects of retrieval support and word-pair difficulty during post-retrieval restudying phase

Comparison of Brain Activation Between Initial Learning and Post-Retrieval Restudying Under No-Retrieval Support Conditions

According to experimental purposes, planned comparisons were conducted using paired-samples t-tests to compare β values for HbO concentration changes between initial learning and post-retrieval restudying phases under no-retrieval support conditions. Results showed that for easy word pairs, brain activation in channel 20 was significantly lower during post-retrieval restudying than during initial learning. For difficult word pairs, post-retrieval restudying showed significantly higher activation than initial learning in channels 12, 14, 17, 18, and 19. Detailed information about significant channels is presented in Table 5 , and brain activation maps are shown in Figure 8a [Figure 8: see original paper].

Table 5 Significantly Activated Channels Comparing Post-Retrieval Restudying and Initial Learning in the No-Retrieval Support Group

Word-Pair Difficulty	Channel	Brodmann Area	Brain Region	t(1,15)	p	95% CI
Easy	20	46	Right dorsolateral prefrontal cortex	-2.34	0.03	[-0.005, -0.001]
Difficult	12	10	Right frontal pole	3.12	0.007	[0.001, 0.007]
	14	45	Right inferior frontal gyrus, pars triangularis	2.89	0.01	[0.001, 0.009]
	17	10	Left frontal pole	3.45	0.003	[0.003, 0.013]
	18	10	Frontal pole	3.21	0.006	[0.002, 0.011]
	19	46	Left dorsolateral prefrontal cortex	2.98	0.009	[0.001, 0.012]

Comparison of Brain Activation Between Initial Learning and Post-Retrieval Restudying Under Retrieval Support Conditions

Planned comparisons were conducted using paired-samples t-tests to compare β values between initial learning and post-retrieval restudying phases under retrieval support conditions. Results showed that for easy word pairs, activation in channels 1, 8, 9, and 16 was significantly lower during post-retrieval restudying than during initial learning. For difficult word pairs, there were no significant differences in brain activation between post-retrieval restudying and initial learning in any channel, ts(1,15) < 2.09, ps > 0.05. Detailed information about significant channels is presented in Table 6 , and brain activation maps are shown in Figure 8b.

Table 6 Significantly Activated Channels Comparing Post-Retrieval Restudying and Initial Learning in the Retrieval Support Group

Word-Pair Difficulty	Channel	Brodmann Area	Brain Region	t(1,15)	p	95% CI
Easy	1	47	Left inferior frontal gyrus	-3.67	<0.001	[-0.017, -0.006]
	8	45	Left inferior frontal gyrus, pars triangularis	-2.98	0.009	[-0.013, -0.003]
	9	47	Right inferior frontal gyrus	-2.76	0.014	[-0.011, -0.002]
	16	46	Left dorsolateral prefrontal cortex	-3.12	0.007	[-0.013, -0.002]

Figure 8 Brain activation maps comparing post-retrieval restudying and initial learning in no-retrieval support (a) and retrieval support (b) groups

3.7 Discussion

When examining the effects of retrieval difficulty and material difficulty on the direct effect of retrieval practice, Experiment 2 results were consistent with Experiment 1, validating the reliability of Experiment 1 findings. Furthermore, when examining the effects on indirect effects, we first observed results consistent with direct effects: the no-retrieval support group showed significantly better final memory performance than the retrieval support group, and easy word pairs showed significantly better final memory performance than difficult word pairs. Second, further analysis revealed that under no-retrieval support conditions, difficult word pairs showed significantly higher final memory performance when given restudying opportunities (compared to the no-restudy group). This result indicates that items that failed retrieval benefit more from subsequent restudying. Particularly for difficult word pairs, using a no-retrieval support strategy for retrieval and receiving restudying opportunities significantly improved memory performance, partially confirming Hypothesis 2.

Regarding the direct effect of retrieval practice, brain imaging results showed that compared with retrieval support conditions, retrieval under no-retrieval support conditions elicited significantly enhanced activation in the superior temporal gyrus. This suggests that the superior temporal gyrus may be associated with retrieval effort during recall. Previous fMRI research has found that superior temporal gyrus activation is related to retrieval search processes, which typically require substantial retrieval effort (Reas & Brewer, 2013). Therefore, the significantly enhanced activation in the superior temporal gyrus in this study supports the notion that more retrieval effort is required in stronger memory search processes, consistent with better final memory test performance under no-retrieval support conditions. Additionally, this study found that retrieving difficult word pairs elicited significantly higher activation levels in multiple prefrontal cortex regions than retrieving easy word pairs, particularly more extensive activation in the dorsolateral prefrontal cortex. Previous neuroimaging research has shown that the dorsolateral prefrontal cortex is sensitive to cognitive load (Pan et al., 2014), with its activation level correlating with increased memory load (Manoach et al., 1997; Toepper et al., 2014). Therefore, participants may have experienced higher cognitive load when retrieving difficult materials, leading to poorer memory performance. This finding suggests that the retrieval effort hypothesis may have boundary conditions; that is, when facing high-difficulty materials, retrieval effort alone is insufficient to ensure good memory outcomes, and cognitive load must also be considered.

Moreover, in exploring the indirect effect of retrieval practice, we found that during the post-retrieval restudying phase, the no-retrieval support group showed significantly stronger prefrontal cortex activation than the retrieval support group. Compared with easy word pairs, difficult word pairs also elicited higher activation levels in the prefrontal cortex. These results indicate that no-retrieval support and difficult word pairs have advantages in promoting the indirect effect of retrieval practice.

Notably, when easy word pairs were restudied after retrieval, activation in relevant prefrontal cortex regions decreased compared with initial learning. In contrast, for difficult word pairs, activation in the frontal pole (associated with memory processing) (Moayedi et al., 2015) and dorsolateral prefrontal cortex (associated with relational memory encoding tasks) (Blumenfeld et al., 2011) significantly increased during restudying under no-retrieval support conditions compared with initial learning. These findings suggest that no-retrieval support strategies can effectively promote subsequent learning of difficult word pairs.

4 General Discussion

This study investigated the effects of retrieval support and word-pair difficulty on the direct and indirect effects of retrieval practice using behavioral experiments and fNIRS technology. Results showed that reducing retrieval support had advantages in enhancing both the direct and indirect effects of retrieval practice, and this advantage was not affected by word-pair difficulty. The effects of word-pair difficulty on the direct and indirect effects of retrieval practice differed: easy word pairs enhanced the direct effect of retrieval practice, but retrieval facilitated subsequent learning of difficult word pairs, particularly under no-retrieval support conditions where the indirect effect of retrieval practice for difficult word pairs was significantly enhanced—a result supported by brain imaging findings. The results support the retrieval effort hypothesis and provide new insights into how retrieval difficulty promotes the retrieval practice effect for difficult materials.

4.1 Effects of Retrieval Support and Word-Pair Difficulty on the Direct Effect of Retrieval Practice

This study found that the no-retrieval support condition significantly enhanced the direct effect of retrieval practice. This result is consistent with Carpenter and DeLosh (2006) and demonstrates that increasing retrieval difficulty prompts learners to invest more retrieval effort, thereby promoting memory retention and supporting the retrieval effort hypothesis and desirable difficulty theory. The fNIRS findings provide neural mechanism evidence for this result and its explanation. Brain imaging data showed that compared with retrieval support conditions, retrieval under no-retrieval support conditions elicited significantly enhanced activation in the superior temporal gyrus region. Previous research has confirmed that enhanced activation in the superior temporal gyrus is associated with greater retrieval effort, particularly during retrieval search processes (Reas & Brewer, 2013). Additionally, recent research has found that activation in temporal regions is associated with better memory performance (Guran et al., 2022; Marin-Garcia et al., 2021). Therefore, when retrieval difficulty is greater, learners may invest more retrieval effort to search for cues associated with target items. This "effort" is crucial for long-term memory retention because it may involve deeper cognitive processing (such as semantic processing and associative memory formation), which provides critical cues for future successful retrieval (Wing et al., 2013).

On the other hand, the study found that easy word pairs showed significantly greater direct benefits from retrieval practice than difficult word pairs. According to cognitive load theory (Sweller, 1988), the complexity of learning materials affects learners' intrinsic cognitive load levels, which in turn influences learning outcomes. Difficult word pairs may impose high intrinsic cognitive load on learners, resulting in lower initial retrieval performance. However, poorer initial retrieval performance is generally not conducive to generating an RPE (Karpicke et al., 2014). fNIRS results showed that retrieving difficult word pairs elicited significantly enhanced activation in the inferior frontal gyrus. Previous research has shown that activation in the inferior frontal gyrus is associated with cognitive effort invested during memory retrieval (van den Broek et al., 2013). This suggests that learners may invest more retrieval effort when retrieving difficult word pairs. However, this retrieval effort did not translate into a memory performance advantage, possibly due to the negative impact of high intrinsic cognitive load. Whelan (2007) proposed that prefrontal cortex activation is highly associated with intrinsic cognitive load. This study observed more extensive activation in the dorsolateral prefrontal cortex when retrieving difficult word pairs than easy word pairs, indicating that difficult word pairs generated higher intrinsic cognitive load, which in turn affected memory performance. Therefore, this study suggests that intrinsic cognitive load level may be a boundary condition for how material difficulty affects the direct effect of retrieval practice. Previous research has suggested that learners invest more retrieval effort when retrieving difficult word pairs, thereby producing a stronger RPE. However, this conclusion may need to consider the impact of intrinsic cognitive load associated with difficult materials.

4.2 Effects of Retrieval Support and Word-Pair Difficulty on the Indirect Effect of Retrieval Practice

This study found that compared with retrieval support conditions, no-retrieval support showed significant advantages in promoting the indirect effect of retrieval practice, and this advantage was not affected by word-pair difficulty, supporting the retrieval effort hypothesis. From a metacognitive perspective, lower initial retrieval success rates under no-retrieval support conditions (as shown in Experiment 1) may prompt learners to form more accurate metacognitive monitoring (Endres et al., 2020). This accurate monitoring helps learners identify and invest more cognitive effort in unmastered content during subsequent learning (Carpenter et al., 2020). Brain imaging results confirmed this finding, showing that the no-retrieval support group exhibited significantly stronger activation in the prefrontal cortex (left and right frontal poles) than the retrieval support group during the post-retrieval restudying phase. The frontal pole, as the largest single structural region in the frontal lobe, plays a key role in higher-order cognition (Burgess et al., 2007) and is associated with cognitive processing of task performance, attention, and episodic memory (Moayedi et al., 2015). Therefore, after experiencing more difficult retrieval, learners invest more attentional resources and encoding effort during subsequent restudying, thereby enhancing the indirect effect of retrieval practice.

On the other hand, unlike direct effects, difficult word pairs showed some advantages over easy word pairs in promoting the indirect effect of retrieval practice. Particularly under no-retrieval support conditions, subsequent restudying of difficult materials helped improve their final memory performance. This finding is consistent with Arnold et al. (2013), which showed that post-retrieval encoding helps promote memory retention, especially for materials that failed initial retrieval. The error correction theory (Grimaldi & Karpicke, 2012) proposes that when learners become aware of discrepancies between their retrieved answers and correct answers, this "error" awareness helps guide systematic adjustments during subsequent learning, thereby improving learning outcomes. Additionally, the search set theory (Grimaldi & Karpicke, 2012; McClelland & Rumelhart, 1985) and mediator effectiveness hypothesis (Pyc & Rawson, 2010) suggest that retrieval processes activate candidate words in the semantic network (search set) that are related to target words, which can serve as mediators or cues for retrieving target words. Even when retrieval fails, candidate words or additional cue words related to target words can still promote re-encoding during subsequent learning, thereby facilitating learning. Combined with the present results, although difficult word pairs had the lowest initial retrieval success rates under no-retrieval support conditions, retrieval promoted encoding during subsequent restudying, helping to enhance the indirect effect of retrieval practice. These findings support theories related to the indirect effect of retrieval practice and reveal the advantageous role of retrieval difficulty in promoting the indirect effect of retrieval practice for difficult word pairs.

Furthermore, during the post-retrieval restudying phase, difficult word pairs elicited higher activation levels than easy word pairs in the frontal pole, inferior frontal gyrus, and dorsolateral prefrontal cortex. Previous research has demonstrated the important role of the frontal pole in attentional control systems (Burgess et al., 2007), and activation in the inferior frontal gyrus is closely related to effort during the encoding phase (Vestergren & Nyberg, 2014). Therefore, retrieval effectively promoted subsequent learning of difficult word pairs, with learners investing more encoding effort and attentional resources during the restudying phase. Additionally, easy word pairs elicited significantly lower prefrontal cortex activity during post-retrieval restudying than during initial learning, whereas difficult word pairs, especially under no-retrieval support conditions, did not show reduced prefrontal cortex activity but instead showed significant enhancement. Royer (1973) noted that during subsequent restudying, successfully retrieved items and failed retrieval items produce opposite effects: encoding of successfully retrieved items may be inhibited, while encoding of failed retrieval items is strengthened. Therefore, for relatively easy-to-retrieve easy word pairs, learners may reduce encoding effort investment during restudying. For difficult word pairs, retrieval may prompt learners to invest more encoding effort during restudying, particularly under no-retrieval support conditions, which may prompt them to strengthen the association between cue and target words (manifested as dorsolateral prefrontal cortex activation), thereby promoting memory retention. This result suggests that difficult retrieval helps learners form more accurate metacognitive monitoring; conversely, relatively easy retrieval may lead to less accurate metacognitive monitoring and insufficient attention to subsequent restudying opportunities.

4.3 Implications and Limitations

This study combined behavioral experiments and brain imaging technology to explore the specific effects of retrieval difficulty and material difficulty on the direct and indirect effects of retrieval practice and their underlying cognitive neural mechanisms. Findings indicate that increasing retrieval difficulty helps enhance both direct and indirect effects of retrieval practice; difficult materials do not have an advantage in promoting the direct effect of retrieval practice, but they help learners form more accurate metacognitive judgments, thereby promoting subsequent learning and ultimately improving learning outcomes (the indirect effect of retrieval practice). These findings have important implications for educational practice, such as encouraging students to actively engage in testing in teaching and recommending low-support retrieval strategies as the preferred testing method. Additionally, for learning difficult materials, the study emphasizes the importance of reviewing (restudying) after testing. These strategies help optimize the direct and indirect effects of retrieval practice in different learning contexts.

However, this study has several limitations. First, regarding word-pair difficulty, this study only included high-difficulty and low-difficulty word pairs, excluding medium-difficulty word pairs. This may limit our understanding of how material difficulty affects the retrieval practice effect. Future research could include medium-difficulty word pairs to more finely explore how material difficulty affects the RPE and further investigate whether intrinsic cognitive load level serves as a boundary condition for how material difficulty affects the RPE. Second, regarding metacognitive judgment measurement, although the study discussed how difficult materials or specific retrieval contexts (such as reduced initial retrieval success rates) help learners form more accurate metacognitive judgments, direct measurement of metacognitive levels to verify this inference was lacking. Future research could combine tasks such as judgments of learning (JOL) (Nelson & Dunlosky, 1991) to monitor learners' metacognitive levels and provide stronger evidence. Finally, regarding brain region exploration, this study focused primarily on the prefrontal cortex. Although this brain region is closely related to the retrieval practice effect, other brain regions such as the temporal or parietal lobes (Marin-Garcia et al., 2021; Zhuang et al., 2022) may also play important roles in retrieval practice. Future research could expand the range of explored brain regions to more comprehensively understand the neural basis of the retrieval practice effect.

5 Conclusion

(1) Increasing retrieval difficulty prompts learners to invest more retrieval effort and increases processing depth during re-encoding (enhanced activation in superior temporal gyrus and prefrontal cortex), effectively enhancing both direct and indirect effects of retrieval practice. (2) Difficult materials do not have an advantage in promoting the direct effect of retrieval practice; however, retrieval can enhance subsequent learning of

difficult materials, and increasing retrieval difficulty helps strengthen the indirect effect of retrieval practice for difficult word pairs.

References

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 940−945.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. Solso (Ed.), *Information Processing and Cognition: The Loyola symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Bjork, E. L. (1992). 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* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Blumenfeld, R. S., Parks, C. M., Yonelinas, A. P., & Ranganath, C. (2011). Putting the pieces together: the role of dorsolateral prefrontal cortex in relational memory encoding. *Journal of Cognitive Neuroscience, 23*(1), 257−265.

Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding: an integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist, 13*(3), 280−291.

Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., Gagnon, L., Boas, D. A., & Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. *Neuroimage, 85*, 181−191.

Burgess, P. W., Gilbert, S. J., & Dumontheil, I. (2007). Function and localization within rostral prefrontal cortex (area 10). *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 887−899.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563–1569.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*, 268−276.

Carpenter, S. K., Endres, T., & Hui, L. (2020). Students' use of retrieval in self-regulated learning: Implications for monitoring and regulating effortful learning experiences. *Educational Psychology Review, 32*(4), 1029−1054.

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language, 92*, 128−141.

Darnai, G., Perlaki, G., Zsidó, A. N., Inhóf, O., Orsi, G., Horváth, R., Nagy, S. A., Lábadi, B., Tényi, D., & Kovács, N. (2019). Internet addiction and functional brain networks: task-related fMRI study. *Scientific Reports, 9*(1), 1−10.

de Lima, M. F. R., Venâncio, S., Feminella, J., & Buratto, L. G. (2020). Does item difficulty affect the magnitude of the retrieval practice effect? An evaluation of the retrieval effort hypothesis. *The Spanish Journal of Psychology, 23*, e31.

Endres, T., Kranzdorf, L., Schneider, V., & Renkl, A. (2020). It matters how to recall–task differences in retrieval practice. *Instructional Science, 48*(6), 699−728.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149−1160.

Greving, S., & Richter, T. (2022). Practicing retrieval in university teaching: short-answer questions are beneficial, whereas multiple-choice questions are not. *Journal of Cognitive Psychology, 34*(5), 657−674.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40*(4), 505−513.

Guran, C. N. A., Deuker, L., Göttlich, M., Axmacher, N., & Bunzeck, N. (2022). Benefit from retrieval practice is linked to temporal and frontal activity in healthy young and older humans. *Cerebral Cortex Communications, 3*(1), tgac009.

Hoshi, Y., Kobayashi, N., & Tamura, M. (2001). Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. *Journal of Applied Physiology, 90*(5), 1657−1662.

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4-5), 528-558.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory, Vol. 2 of Learning and memory: A comprehensive reference* (J. H. Byrne, Series Ed.) (pp. 487-514). Oxford: Academic Press.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *The psychology of learning and motivation* (vol. 61, pp. 237–284). San Diego, CA: Elsevier Academic Press.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966−968.

Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology, 115*, 101237.

Kovelman, I., Shalinsky, M. H., Berens, M. S., & Petitto, L.-A. (2008). Shining new light on the brain's "bilingual signature": a functional near infrared spectroscopy investigation of semantic processing. *Neuroimage, 39*(3), 1457−1471.

Ma, X. F., Li, Z. Y., & Zhou, A. B. (2022). Can Retrieval Practice Promote the Learning of Procedural Knowledge? ——Testing Based on Prompt Retrieval. *Journal of Psychological Science, 45*(4), 849−855. [马小凤, 李增艳, 周爱保. (2022). 提取练习能够促进程序性知识的学习吗?——基于有提示提取的检验. 心理科学, 45(4), 849−855.]

Ma, X. F., Zhou, A. B., & Yang, X. E. (2017). Cue Strength: The Important Variables of The Examination of Retrieval Practice Effect Mechanism. *Psychological Development and Education, 33*(3), 313−320. [马小凤, 周爱保, 杨小娥. (2017). 线索强度:检验提取练习效应内部机制的重要变量. 心理发展与教育, 33(3), 313−320.]

Manoach, D. S., Schlaug, G., Siewert, B., Darby, D. G., Bly, B. M., Benfield, A., … & Warach, S. (1997). Prefrontal cortex fMRI signal changes are correlated with working memory load. *Neuroreport, 8*(2), 545−549.

Marin-Garcia, E., Mattfeld, A. T., & Gabrieli, J. D. (2021). Neural correlates of long-term memory enhancement following retrieval practice. *Frontiers in Human Neuroscience, 15*, 584560.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology. General, 114*, 159–188.

McDermott, K. B. (2021). Practicing retrieval facilitates learning. *Annual Review of Psychology, 72*, 609−633.

Moayedi, M., Salomons, T. V., Dunlop, K. A., Downar, J., & Davis, K. D. (2015). Connectivity-based parcellation of the human frontal polar cortex. *Brain Structure and Function, 220*, 2603−2616.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science, 2*(4), 267–270.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402−407.

Nelson, S. M., Arnold, K. M., Gilmore, A. W., & McDermott, K. B. (2013). Neural signatures of test-potentiated learning in parietal cortex. *Journal of Neuroscience, 33*(29), 11754−11762.

Noble, W. S. (2009). How does multiple testing correction work? *Nature Biotechnology, 27*, 1135−1137.

Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., Oda, I., Isobe, S., Suzuki, T., & Kohyama, K. (2004). Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10–20 system oriented for transcranial functional brain mapping. *Neuroimage, 21*(1), 99—111.

Pan, J. J., Jiao, X. J., Jiang, J., Xu, F. G., & Yang, H. J. (2014). Mental Workload Assessment Based on Functional Near-Infrared Spectroscopy. *Acta Optica Sinica, 34*(11), 344—349. [潘津津, 焦学军, 姜劲, 徐凤刚, 杨涵钧. (2014). 利用功能性近红外光谱成像方法评估脑力负荷. 光学学报, 34(11), 344—349.]

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences, 1464*(1), 5—29.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*(6002), 335—335.

Pyc, M. A., & Rawson, K. A. (2012). Why is test–restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 737—746.

Reas, E. T., & Brewer, J. B. (2013). Retrieval search and strength evoke dissociable brain activity during episodic memory recall. *Journal of Cognitive Neuroscience, 25*(2), 219—233.

Roediger III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181—210.

Royer, J. M. (1973). Memory effects for test-like-events during acquisition of foreign language vocabulary. *Psychological Reports, 32*(1), 195—198.

Shu, H., & Zhang, Y. X. (2008). *Methods in psychological science: Experimental design and data analysis.* Beijing: People's Education Press. [舒华, 张亚旭. (2008). 心理学研究方法：实验设计和数据分析. 北京：人民教育出版社.]

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory, 22*(7), 784—802.

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

Toepper, M., Gebhardt, H., Bauer, E., Haberkamp, A., Beblo, T., Gallhofer, B., … & Sammer, G. (2014). The impact of age on load-related dorsolateral prefrontal cortex activation. *Frontiers in Aging Neuroscience, 6*, 9.

Van den Broek, G. S., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage, 78*, 94—102.

Vaughn, K. E., Rawson, K. A., & Pyc, M. A. (2013). Repeated retrieval practice and item difficulty: Does criterion learning eliminate item difficulty effects? *Psychonomic Bulletin & Review, 20*(6), 1239−1245.

Vestergren, P., & Nyberg, L. (2014). Testing alters brain activity during subsequent restudy: evidence for test-potentiated encoding. *Trends in Neuroscience and Education, 3*(2), 69−80.

Wan, N., Hancock, A. S., Moon, T. K., & Gillam, R. B. (2018). A functional near-infrared spectroscopic investigation of speech production during reading. *Human Brain Mapping, 39*(3), 1428−1437.

Wang, Z. X., & Zhang, J. K. (2017). The mechanism of the retrieval practice effect: From encoding, retrieval to integration of both. *Journal of Psychological Science, 25*(9), 1512−1520. [王植洵, 张锦坤. (2017). 提取练习效应的产生机制：编码、提取的整合演变. 心理科学进展, 25(9), 1512−1520.]

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*(1), 1−12.

Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. *Neuropsychologia, 51*(12), 2360−2370.

Wissman, K. T., & Rawson, K. A. (2018). Test-potentiated learning: Three independent replications, a disconfirmed hypothesis, and an unexpected boundary condition. *Memory, 26*(4), 406−414.

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: a network visualization tool for human brain connectomics. *Plos One, 8*(7), e68910.

Yang, H. B., Liu, H. J., Zhang, P., & Li, L. (2019). The role of masking stimulation in target recognition processing: Evidence from fNIRS. *Acta Psychologica Sinica, 51*(11), 1187−1197. [杨海波, 刘和珺, 章鹏, 李量. (2019). 掩蔽刺激对目标识别加工的作用：来自 fNIRS 的证据. 心理学报, 51(11), 1187−1197.]

Yang, L. X., Zhang, J. K., Li, D. J., & Zhang, L. J. (2022). The Longer the Retrieval Process, The Better the Memory Retention? The Moderating Effect of Material Difficulty. *Journal of Psychological Science, 45*(3), 567−573. [杨丽娴, 张锦坤, 李冬静, 张俐娟. (2022). 提取过程越久记忆保持越好吗？材料难度的调节作用. 心理科学, 45(3), 567−573.]

Ye, Z., Shi, L., Li, A., Chen, C., & Xue, G. (2020). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *ELife, 9*, e57023.

Zhang, H., Hou, S., Wang, H. M., Lian, Y. X., & Yang, H. B. (2020). Socially shared retrieval-induced forgetting in a naturalistic collaborative retrieval situation. *Acta Psychologica Sinica, 52*(6), 716−729. [张环, 侯双, 王海曼, 廉宇煊, 杨海波. (2020). 他人在场条件下的社会分享型提取诱发遗忘. 心理学报, 52(6), 716−729.]

Zhang, J. K., & Zhang, L. J. (2020). The Effects of Encoding and Retrieval Duration on Retrieval Practice Effect. *Journal of Psychological Science, 43*(4), 785–792. [张锦坤, 张俐娟. (2020). 编码与提取时长对提取练习效应的影响. 心理科学, 43(4), 785–792.]

Zhuang, L., Wang, J., Xiong, B., Bian, C., Hao, L., Bayley, P. J., & Qin, S. (2022). Rapid neural reorganization during retrieval practice predicts subsequent long-term retention and false memory. *Nature Human Behaviour, 6*(1), 134–145.

Note: Figure translations are in progress. See original paper for figures.