



Improving the Use of Retrieval Practice for Both Easy and Difficult Materials: The Effect of an Instructional Intervention

Tian Fan¹ · Luotong Hui^{2,3} · Liang Luo⁴ · Anique B. H. de Bruin²

Accepted: 27 August 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Recent research has suggested that students prefer restudying over retrieval practice when learning difficult materials, despite the latter being a more effective learning strategy. The current study investigated whether an instructional intervention can improve the use of retrieval practice for both easy and difficult materials. In Experiment 1, after initial learning of each item, participants rated their perceived mental effort (PME) and judgment of learning (JOL) for each item. Then, participants chose whether to restudy or take retrieval practice for that item. The results showed that participants chose to take retrieval practice less frequently for difficult items compared to easy ones. Furthermore, participants' ratings of PME and JOL sequentially mediated the relationship between item difficulty and their learning strategy choices. Specifically, difficult items resulted in higher levels of PME, which in turn led to lower JOL, ultimately reducing the likelihood of choosing retrieval practice. In Experiment 2, half of the participants received an instructional intervention, which revealed that while students prefer restudying for difficult items, retrieval practice benefits both easy and difficult items in long-term retention. The remaining half did not receive such intervention and were designated as the control group. The results indicated that, compared to the control group, students who received the intervention increased the odds of choosing retrieval practice for both types of materials after the intervention. The findings of this study suggest that students can be supported to use retrieval practice regardless of item difficulty.

Keywords Retrieval practice · Item difficulty · Intervention · Self-regulated learning · Effective learning strategy

Extended author information available on the last page of the article

Published online: 02 October 2024

Springer

Introduction

In the wake of advancements in science and technology, there are numerous options for accessing educational materials. Consequently, an increasing number of learning activities occur outside of traditional classroom settings, which requires students to manage learning on their own. An important and timely research topic pertains to understanding students' self-regulated learning (Bjork et al., 2013). In authentic learning circumstances, the content often varies in difficulty. For example, when students memorize vocabulary for a new language, they must memorize not only the easy words but also the difficult words. How do students make decisions about how to learn materials with varying difficulty? Do students employ effective learning strategies for both easy and difficult materials? The present study focuses on students' utilization of a specific strategy known as retrieval practice (or self-testing) in the context of self-regulated learning.

The subsequent section offers a brief overview of how retrieval practice works as an effective learning strategy for both easy and difficult materials. This is followed by recent empirical studies indicating that students typically engage in less retrieval practice for difficult materials compared to easy ones. Based on this overview, the current study underscores the critical necessity to assist students in increasing their use of retrieval practice regardless of item difficulty.

Retrieval Practice as an Effective Learning Strategy for Both Easy and Difficult Materials

Hundreds of studies spanning more than a century of research have demonstrated that retrieval practice is one of the most powerful ways to enhance long-term retention (Dunlosky et al., 2013). The phenomenon that retrieving information from memory consolidates long-term retention better than restudying the same information has been termed the *testing effect* or *test-enhanced learning* (Roediger & Butler, 2011; Roediger & Karpicke, 2006a). The testing effect is a robust phenomenon across a wide variety of learning materials, including single words (e.g., Zaromb & Roediger, 2010), paired associates (Carpenter, 2009), and prose passages (e.g., Roediger & Karpicke, 2006b). Moreover, the testing effect has been reliably found not only in the laboratory but also in real classrooms and across educational levels (for recent reviews, see Adesope et al., 2017; Carpenter et al., 2022; Rowland, 2014; Yang et al., 2021).

Previous studies have demonstrated that retrieval practice benefits both easy and difficult items in the long term (de Jonge & Tabbers, 2013). For example, de Jonge and Tabbers (2013) asked participants to learn both easy and difficult word pairs under conditions of repeated study and repeated tests (without feedback). Results showed that on a 1-week retention test, the repeated testing condition outperformed the repeated study condition for both types of materials. Another illustration comes from Minear et al. (2018), in which participants studied easy and difficult Swahili-English word pairs and repeatedly studied half the pairs and took retrieval practice, with feedback, for the remaining half. The delayed test initiated 2 days later

suggested that the tested items were recalled more than the repeatedly studied items for both easy and difficult items. Additionally, the benefits of retrieval practice over restudying did not differ significantly between easy and difficult items.

Moreover, recent studies have indicated that the inclusion of feedback during retrieval practice can enhance the testing effect (Rowland, 2014). In the absence of feedback, the retrievability of study materials becomes crucial for the positive testing effect to occur (Jang et al., 2011). A review conducted by Rowland (2014) revealed that in studies lacking feedback, initial test performance of less than or equal to 50%, indicating the learning materials were difficult, did not produce a reliable testing effect ($g=0.03$). Conversely, studies incorporating feedback exhibited the numerically largest testing effects ($g=0.73$), irrespective of retrieval success. These findings suggest that combining retrieval practice with feedback can augment the effectiveness of retrieval practice, particularly when learning challenging materials. To foreshadow, the current study provided participants with feedback about the correct answers during retrieval practice.

Carpenter et al. (2017) investigated students' self-regulated use of retrieval practice (vs. restudy) for online study materials and its relationship to exam performance. Students in an introductory biology course were provided with optional online review questions that could be accessed through retrieval practice (i.e., answering questions before receiving feedback) or through restudying (i.e., providing questions and correct answers up-front). The results indicated that the highest-performing students on exams were those who exclusively used retrieval practice for all materials, followed by those who employed both retrieval practice and restudy strategies, while the lowest performance was observed in those who solely relied on the restudy strategy. Moreover, there was a generally positive correlation between the amount of retrieval practice completed and exam performance. Such finding suggested that the more students engage in retrieval practice, the better their learning outcomes tend to be.

In summary, a wealth of research in cognitive and educational psychology underscores that retrieval practice is an effective learning strategy in enhancing long-term retention on both easy and difficult materials. Therefore, students should be encouraged to optimize the use of retrieval practice for both types of materials during self-regulated learning.

Students' Use of Retrieval Practice for Easy and Difficult Materials

In authentic learning contexts, students encounter both easy and difficult content, making it crucial to explore how they regulate their use of retrieval practice across varying difficulty levels. Previous research has addressed this issue, indicating that students tend to favor restudying over retrieval practice for difficult learning materials.

For example, Toppino et al. (2018) conducted a study wherein college students studied easy and difficult word pairs (i.e., associated vs. unrelated pairs). Participants were given the option to restudy the pair, take retrieval practice, or discontinue further study after initial learning. The study also manipulated spacing

between-subjects: retrieval practice or restudy occurred either after the presentation of two other pairs (short-lag condition) or after all pairs had been presented (long-lag condition). Item difficulty was indexed by judgment of learning (JOL), with low, medium, and high JOL ratings indicating varying levels of difficulty. Results showed that, regardless of whether practice tests were followed by feedback, participants significantly preferred testing over restudying for high JOL items (easy items) in both short-lag and long-lag conditions. Conversely, participants opted to restudy significantly more often than to test themselves for low JOL items (difficult items) in long-lag conditions, indicating differential regulation of retrieval practice for easy and difficult items.

Similarly, Tullis et al. (2018) investigated whether students' use of retrieval practice depended on item difficulty. They asked participants to learn associated (easy) and unassociated (difficult) word pairs and then choose to restudy or test for each pair. Across five experiments, participants consistently chose to test themselves more frequently on associated pairs than on unassociated pairs, regardless of (1) whether participants were required to select precisely half of the items to restudy and the other half to self-test or unlimitedly chose to self-test as many word pairs as they wanted; (2) whether feedback was provided during retrieval practice or not; and (3) whether participants were college students or were recruited online through Amazon Mechanical Turk.

In a recent study by Badali et al. (2022), undergraduate students learned normatively easy and difficult Lithuanian-English word pair translations. After an initial study trial, students in self-regulated learning groups decided whether to restudy the item, take a practice test, or drop the item from further practice. Feedback about the correct English word was provided after retrieval practice attempts. Unlike the prior two studies, participants in this study could choose to restudy or self-test the items multiple times before dropping them, and they could make new choices for items each time they were presented. Results indicated that participants restudied a significantly higher proportion of difficult items than easy ones and tested a significantly lower proportion of difficult items than easy ones during initial learning strategy choices.

The above-mentioned studies collectively suggested that students selectively employ retrieval practice based on item difficulty, using it less frequently for difficult items compared to easy ones. Building upon these studies, the current study aimed to provide new insights into how students regulate their use of retrieval practice when learning highly educationally relevant materials, such as human anatomical image-name pairs. Freshmen majoring in health sciences often need to memorize human anatomical structures with varying difficulty, making it essential to understand how they initiate effective learning strategies on easy and difficult materials to optimize their learning. To foreshadow, the current study was conducted online, fostering a less supervised and more authentic self-regulated learning environment. A pilot experiment was conducted to determine the item difficulty of the anatomical image-name pairs (see Online Supplementary Materials: Appendix A for details). Item difficulty was defined as the percentage of participants who accurately recalled the name when given images as cues, with

items scoring in the upper one-third selected as easy items, while items scoring in the lower one-third were selected as difficult items.

Factors Influencing Students' Use of Retrieval Practice for Easy and Difficult Materials

Retrieval practice is considered a form of *desirable difficulty*, as described by Bjork and Bjork (2011). This concept suggests that although retrieval practice may be experienced as effortful, it ultimately enhances long-term retention. However, students often fail to recognize the direct benefits of retrieval practice for memory and may only appreciate its indirect effects, such as using it solely as an assessment tool at the end of their learning period (e.g., Rivers, 2021; Tullis & Maddox, 2020). The delayed and long-term nature of the benefits of retrieval practice for memory are often less immediately apparent to students, leading to negative perceptions of retrieval practice, such as high perceived mental effort (PME) and/or feelings of learning less. Consequently, students may exhibit resistance to engaging in retrieval practice and prefer the alternative strategy of restudying.

Recently, a study by Kirk-Johnson et al. (2019) proposed a misinterpreted-effort hypothesis to explain this phenomenon. Participants were asked to employ both retrieval practice and restudying learning strategies, and then report their perceived effort and learning *associated with each strategy* through a questionnaire. Subsequently, participants were asked to choose one of the two strategies for future learning. The findings indicated that the more mentally effortful participants perceived retrieval practice relative to restudying, the less effective they perceived retrieval practice to be for learning, and in turn, the less likely they were to choose it over restudying. This study suggests that students felt retrieval practice as an effort-intensive strategy.

In self-regulated learning, students need a clear understanding of the materials they are learning (e.g., easy or difficult materials) before making decisions about how to study. While strategy-level experiences influence students' decisions (as discussed above), task-level experiences, such as the PME and perceived learning associated with the studied items, also play a crucial role in decisions about how to learn the items. A recent study by Hui et al. (2022) tested this misinterpreted-effort hypothesis at the learning task level. In this study, students learned image-name pairs, and their perceived effort and perceived learning for each pair were measured during learning. Their decisions for future learning of each pair (i.e., retrieval practice *vs.* restudy) were recorded. Initially, when students did not receive any performance feedback, students' decisions about learning strategy were directly influenced by their PME about the pairs, while perceived learning did not play a significant role. This result suggested that prior to receiving performance feedback, students primarily relied on PME associated with the pairs as a direct indicator for their learning strategy decisions. In other words, students believed that learning items requiring low mental effort allowed them to adopt an effort-intensive strategy like retrieval practice, while items demanding high mental effort called for a strategy requiring less mental effort, such as restudying.

Based on the above studies, it appears that when making decisions about whether to restudy or take retrieval practice for each easy or difficult item, students rely on both strategy-level and task-level experiences. Seufert's (2018, 2020) model of *self-regulation as a function of resources and perceived cognitive load* suggested that students may balance between task-level experiences (e.g., PME and perceived learning associated with the easy or difficult materials) and strategy-level experiences (e.g., the demands of the strategy itself) when deciding whether to restudy or engage in retrieval practice for each easy or difficult item.

Task difficulty is a crucial parameter of Seufert's model, which proposes that resources and cognitive load are two important mediators influencing the relationship between task difficulty and self-regulatory activities. One important assumption of Seufert's model is that learners' resources decrease as task-difficulty increases, which in turn mediates the effect on self-regulatory activities. Specifically, learning items of varying difficulty and choosing a strategy (retrieval practice vs. restudy) at the same time require working memory capacity. Learning difficult items demands considerable effort and occupies a significant portion of working memory capacity. With insufficient capacity left, learners may be unable to effectively engage in self-regulatory processes. As a consequence, learners may resort to less effective but less effort-intensive strategy (e.g., restudying).

Based on the above studies, learning easy items typically requires low mental effort from participants, often resulting in a perception of high learning. Since these task-level experiences do not impose significant cognitive load on students (i.e., students' resources are sufficient), they are more likely to opt for the effort-intensive strategy of retrieval practice. Conversely, learning difficult materials may require more mental effort compared to easy items, leading to lower perceived learning. As these task-level experiences are already demanding (i.e., students' resources are insufficient), students may opt for the alternative strategy—restudying—which is less effort-intensive. This hypothesized process can be investigated using a serial mediation model: examining whether the influence of item difficulty (independent variable) on students' learning strategy decisions (outcome variable, retrieval practice vs. restudy) is sequentially mediated by their PME and JOL associated with the materials. As students tend to differentially use retrieval practice based on item difficulty, it is crucial to understand the factors influencing students' decision-making processes.

Improving the Use of Retrieval Practice for Both Easy and Difficult Materials

Previous studies have shed light on the efficacy of interventions aimed at enhancing students' utilization of retrieval practice. A recent review by Carpenter (2023) comprehensively synthesizes emerging literature on interventions tailored to bolster students' inclination toward employing retrieval practice. For instance, some studies advocate for multifaceted approaches to augment students' engagement in retrieval practice (Biwer et al., 2020; Broeren et al., 2021). Biwer et al. (2020) developed a learning strategy intervention program called "Study Smart" with the objective of fostering awareness, reflection, and

implementation of effective learning strategies, including retrieval practice. The program comprised three 2-hour sessions administered to first- and second-year undergraduate students. Following the intervention, students reported an increase adoption of effective learning strategies including retrieval practice. Additionally, there was a significant increase in students' knowledge of such strategies. Subsequently, Biwer et al. (2022) further investigated the effect of the "Study Smart" program on improving students' academic performance. In this investigation, all first-year pharmacology students attended the "Study Smart" program in their first weeks. The 20% lowest performing students in the first midterm exam received further support regarding their use of learning strategies. Results showed that students in the Study Smart cohort exhibited notable improvement in their academic performance across exams compared to the control cohort. Furthermore, the differences in the final exam test scores between the top, middle, and bottom ranks were reduced in the Study Smart cohort compared to the control cohort. This study underscores that the "Study Smart" program, coupled with a remediation track for underperforming students, can effectively enhance students' study habits and academic achievements.

Some laboratory-based studies designed interventions in a relatively straightforward manner, which can be easily adopted to foster students' use of retrieval practice. These interventions involved providing students with feedback on the beneficial outcomes of retrieval practice (Ariel & Karpicke, 2018; Hui et al., 2021). For example, Hui et al. (2021) enabled students to experience the benefits of retrieval practice by showing and comparing their test performance after studying human anatomical structures through restudying and retrieval practice. Provision of feedback on actual learning outcomes is imperative to rectify students' misconceptions regarding effort and learning efficacy, as it underscores the long-term benefits of the challenging learning strategy—retrieval practice. Following such feedback, students who had experienced the testing effect were more inclined to use retrieval practice compared to those who had not. Hence, providing feedback on the positive consequences of retrieval practice emerges as an effective way in fostering its utilization among students.

It should be noted that prior studies on interventions targeting retrieval practice did not show the benefits of retrieval practice *separately* for easy and difficult items. Since students tend to employ retrieval practice differentially based on item difficulty, using more restudying (and less retrieval practice) for difficult items compared to easy ones, additional guidance may be necessary to encourage students to engage in retrieval practice for difficult items. Anticipating this need, the intervention in the present study explicitly addresses students' selective use of learning strategy based on item difficulty. Furthermore, the intervention specifically illustrates that compared with restudying, retrieval practice benefits both easy and difficult items in long-term retention. Such intervention may alleviate students' potential reservations about employing retrieval practice for challenging materials and increase the use of retrieval practice for both types of materials.

Overview of the Current Study

The current study aimed to address three primary questions across two experiments. In Experiment 1, we investigated (1) how students regulate their use of retrieval practice for easy and difficult anatomical image-name pairs and (2) how task-level experiences, more specifically, PME and JOL associated with the items, influence students' learning strategy choices for easy and difficult items. Participants rated their PME and JOL after initial learning of each easy or difficult item and subsequently selected whether to restudy or engage in retrieval practice for future learning. In Experiment 2, we explored (3) whether an instructional intervention on retrieval practice improves the use of retrieval practice for both types of materials. Participants were divided into two groups, with one group receiving the instructional intervention and the other receiving unrelated reading materials. The intervention comprised two steps. Firstly, it highlighted to students that individuals tend to choose restudy more frequently for difficult items than for easy ones. Secondly, the intervention presents test performance outcomes after studying both types of materials through restudying and retrieval practice, showing students that retrieval practice, rather than restudying, benefits both easy and difficult items in long-term retention. Unlike previous studies, the current investigation separately demonstrated to students the learning benefits of retrieval practice for both easy and difficult materials. To our knowledge, the present study is the first to investigate whether such an intervention can improve the use of retrieval practice for both easy and difficult materials.

Experiment 1

In Experiment 1, participants rated their PME and JOL for each easy or difficult item after initial learning. Subsequently, they decided whether to restudy or engage in retrieval practice for that item over three additional learning rounds. We hypothesized that:

Hypothesis 1: Participants choose to take retrieval practice on a higher percentage of easy items compared to difficult items.

Hypothesis 2: Participants' PME and JOL play serial mediating roles in the relationship between item difficulty and learning strategy choices (i.e., difficult items (compared to easy ones) \rightarrow greater PME \rightarrow lower JOL \rightarrow decreased likelihood of choosing retrieval practice). In addition, Experiment 1 explored how students appreciate the effectiveness of retrieval practice and restudy at different time points. Since this aspect constitutes an exploratory investigation, no *a priori* hypotheses were formulated.

Method

Participants

A pilot experiment with 24 participants observed Cohen's $d=0.51$ for the difference in the percentage of easy and difficult pairs that participants chose to study further using retrieval practice ($M_{\text{easy}}=60.21\%$, $SD_{\text{easy}}=36.99\%$; $M_{\text{difficult}}=55.42\%$, $SD_{\text{difficult}}=36.53\%$). Based on this effect size, a power analysis was conducted using G*Power (Faul et al., 2007), indicating that 44 participants were required to detect a significant (one-tailed, $\alpha=0.05$) difference at 95% power. To be more conservative about the data quality of online experiments, we decided to increase the sample size to 70.

Seventy participants (51 females; age: $M=21.79$ years, $SD=2.13$) were recruited from Prolific (<https://www.prolific.com/>) for monetary compensation (£6.50; the top 50% highest scoring participants on the memory test received an additional bonus of £1.50). All participants spoke English as a first language, did not study medical or biological subjects, and did not have reading disorders. Educational levels varied, with participants falling into the following categories: high school graduate (18.6%), some college but no degree (42.9%), 2-year associate degree in college (4.3%), 4-year Bachelor's degree in college (32.9%), and Master's degree (1.4%). Approximately 40% of participants had never learned human anatomy before, while the others learned some before or in university. The majority of participants (87.1%) had never learned Latin while 94.3% of them had never learned ancient Greek, indicating limited familiarity with the anatomical names which are often closely related to Latin or ancient Greek. All procedures were approved by the Ethics Review Committee of the Faculty of Health, Medicine, and Life sciences at Maastricht University (approval number: FHML-REC/2021/121), and informed consent was obtained from all participants.

Materials

Memory Tasks Forty-eight image-name pairs of human anatomical structures were used for Experiment 1. These pairs consisted of images depicting anatomical structures accompanied by their corresponding names. Eight pairs were used for practice and were subsequently excluded from data analyses. The remaining 40 pairs were used for the first learning task (see Fig. 1), half of which were easy items and the others were difficult. A pilot experiment was conducted to select easy and difficult image-name pairs (see Online Supplementary Materials: Appendix A for details). The 40 anatomical image-name pairs were organized into five units (units 1–5), each comprising eight pairs (four easy and four difficult pairs). Compared to a smaller unit size (e.g., four pairs), where participants could readily retrieve the name of the structure after a short delay (e.g., at most three pairs), the larger size of eight pairs per unit necessitated increased and desirable retrieval effort. Additionally, presenting 40 pairs in an intermixed manner places greater demands on working memory

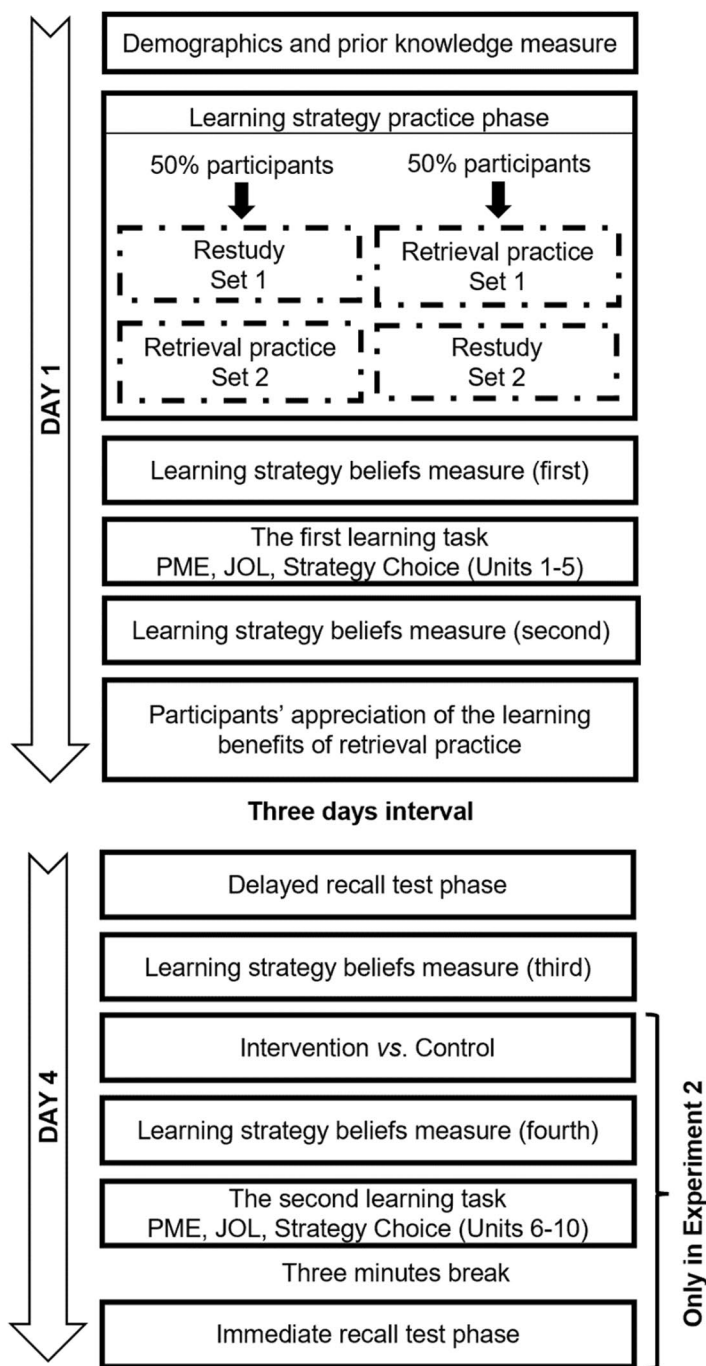


Fig. 1 Procedures of Experiments 1 and 2. See Fig. 2 for details of the procedure for the first and second learning tasks, during which participants rated their perceived mental effort (PME), made judgment of learning (JOL), and chose their learning strategy

capacity compared to organizing pairs into units. Therefore, organizing pairs into units and with eight pairs per unit may represent a moderated design for a learning task.

Measures

Learning Strategy Beliefs Participants' beliefs regarding retrieval practice and restudy were assessed through two questions: "How effective is restudying (or self-testing) in helping you to memorize the anatomical image-name pairs from 1 (extremely ineffective) to 7 (extremely effective)?" The two questions appeared in a random order. These two questions on learning strategy beliefs were administered multiple times throughout the procedure, as depicted in the "Learning strategy beliefs measure" boxes in Fig. 1.

Perceived Mental Effort (PME) The PME was measured by asking participants two questions for each pair. The first one was: "How much mental effort did this image-name pair require from you to learn from 1 (very, very low mental effort) to 9 (very, very high mental effort)?" (Paas, 1992). The second one was: "How difficult was it for you to learn this image-name pair from 0 (not difficult at all) to 10 (extremely difficult)?" These two aspects together provided an estimation of overall mental effort (Hui et al., 2022; van Gog & Paas, 2008). To foreshadow, the correlation between the two questions was .89 ($p < .001$) in Experiment 1, .89 ($p < .001$) in the first learning task of Experiment 2, and .86 ($p < .001$) in the second learning task of Experiment 2. In the current study, PME was calculated as the average of the two measures. Specifically, the value of participants' ratings of the required mental effort was converted from 1–9 to 0–10, and was added to the value of the perceived item difficulty and divided by 2 (Hui et al., 2022).

Judgment of Learning (JOL) Participants were asked to estimate the likelihood that they would remember the image-name pair after 3 days, by the question: "How likely are you to correctly recall the name of this anatomical image after three days from 0 (sure I will not recall it correctly) to 100 (sure I will recall it correctly)?" (Koriat, 1997).

Learning Strategy Choice Participants' learning strategy choices were assessed through one binary question: "You have studied this image-name pair once. Please choose a learning strategy for further study. You will study this image-name pair three more times with your chosen strategy. You want to: A. continue with restudying, or B. continue with self-testing?". The order of the two options was randomized.

Participants' Appreciation of the Learning Benefits of Retrieval Practice Two questions were adopted from previous studies (Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007; Yan et al., 2014). The first question aimed to investigate why students choose retrieval practice, allowing participants to select multiple options. The second question probed participants' beliefs about the usefulness of retrieval practice.

Instead of asking participants to choose one statement they most agreed with, they were asked to rate their agreement with four statements on a scale from 1 (strongly disagree) to 5 (strongly agree). Since this aspect was not our primary focus, it was reported in the supplementary materials (see Online Supplementary Materials: Appendix B).

Procedure

The current study was conducted using the Qualtrics survey software (<https://www.qualtrics.com/>). Participants were instructed to allocate approximately 60 min to the experiment and complete it without interruption. The experiment consisted of two parts (Day 1 and Day 4), separated by a 3-day interval. The procedure is depicted in Fig. 1.

At the beginning of the first day, participants answered questions about demographics, prior knowledge of anatomy, and knowledge of Greek and Latin. Participants then started with a learning strategy practice phase, during which they familiarized themselves with restudy and retrieval practice strategies using two sets of image-name pairs. Half of the participants began with restudy set 1 (four pairs) followed by retrieval practice on set 2 (four pairs), while the other half started with retrieval practice on set 1 and restudy set 2. Each set was presented for four rounds. During the first round of study, participants studied each pair for 8 s one after another. During the next three rounds of study, for the restudy strategy, participants studied the anatomical image along with its name for 8 s. For the retrieval practice strategy, participants were shown the anatomical image and the first letter of its name and were given 6 s to recall and type the full name into an entry box. The feedback about the correct name was then provided to the participants for 2 s. For example, “*Your answer is correct (or wrong). The correct name of this image is Spleen.*” Following the second learning round, participants completed a 15-s distractor activity that required them to write down as many countries across different regions in the world as possible.

Upon completion of the learning strategy practice phase, participants’ beliefs about retrieval practice and restudy were measured for the first time (see the questions titled “learning strategy beliefs” in the “Measures” section). Subsequently, participants started the learning task, in which they studied five units of image-name pairs one after another. The order of presenting these five units was randomized. The procedure of the learning task is presented in Fig. 2. During the first round of learning, after an 8-s presentation of each image-name pair, participants were asked to answer two questions about PME and one question about JOL before choosing a learning strategy to use for further study. Different rating scales (0–10, 1–9, and 0–100%, respectively) and formats (vertical and horizontal) were used to strengthen the distinction between the rating questions (Koriat et al., 2014). Although the first three questions appeared randomly, the last question was always the learning strategy choice (see the “Measures” section). During the next three learning rounds, participants used their chosen strategy (restudy or retrieval practice) to study each

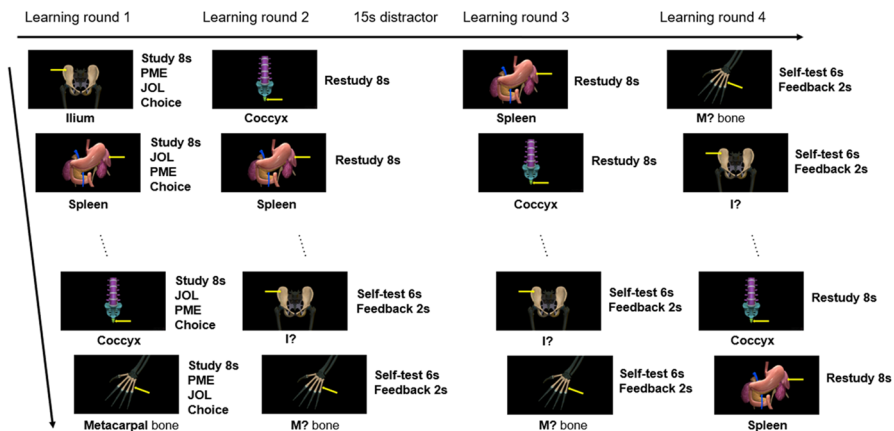


Fig. 2 Display of the four learning rounds in each unit during the learning task. For each round, the eight pairs were presented in a random order. All images were obtained from www.anatomylearning.com. PME=perceived mental effort, JOL=judgment of learning

image-name pair as they did during the learning strategy practice phase. This process continued until participants completed all five units.¹ After that, participants' beliefs about retrieval practice and restudy were measured for the second time. Participants were then asked two questions about their appreciation of the learning benefits of retrieval practice (see the “Measures” section and Online Supplementary Materials: Appendix B). At the end of Day 1, participants were prompted to report any internet connection problems or disturbances encountered during the experiment.

After a 3-day interval, participants took a delayed test on the 40 image-name pairs that they had memorized on Day 1. During the test, each image was presented alongside the first letter of its corresponding name, and participants were required to recall the name from memory and type it into an entry box. There was no time limitation to type the answer, and no feedback was provided on the response. Participants proceeded to the next item by pressing a button after entering their response for each item. After the delayed recall test, participants' learning strategy beliefs were measured for the third time. This repeated measurement of learning strategy beliefs aimed to investigate any changes in participants' beliefs about the two learning strategies over time. Participants' recall performance scores were then shown to them (e.g., Congratulations! You correctly answered X of the 40 image-name pairs!). At the end of Day 4, participants were prompted to report any internet connection problems, disturbances experienced, or assistance received from external sources during the test.

¹ Participants received a 3-min break after finishing two units of studying in which they watched a cartoon.

Results

All participants passed the attention checks during the experiment. Analyses on the delayed recall performance and learning strategy beliefs excluded data from seven participants who did not return to complete these sections on Day 4. Below, we presented the results relevant to our main research questions. Additional results of potential interest to readers were provided in the Online Supplementary Materials.

Delayed Recall Performance

We calculated the mean percentage of items participants correctly recalled on the delayed test as functions of item difficulty (easy vs. difficult) and the learning strategy participants chose to use during the learning phase (see Table 1). A 2 (item difficulty: easy vs. difficult) \times 2 (strategy: restudy vs. retrieval practice) within-subjects repeated-measures analysis of variance (ANOVA) was conducted on the delayed recall performance. The main effect of item difficulty was significant, $F(1, 27) = 64.35$, $p < .001$, $\eta_p^2 = 0.70$, indicating that the delayed recall performance for easy items was higher than for difficult items. The main effect of strategy was significant, $F(1, 27) = 39.89$, $p < .001$, $\eta_p^2 = 0.60$, indicating that recall performance after retrieval practice was higher than after restudy. Since participants' learning strategy choices were self-regulated, this main effect may be confounded with item difficulty (e.g., participants may have chosen retrieval practice on relatively easier items). The interaction between item difficulty and strategy was not significant, $F(1, 27) = 0.01$, $p = .93$, $\eta_p^2 < 0.001$.

Retrieval Practice Choice

To address the first research question (i.e., How do students regulate their use of retrieval practice for easy and difficult anatomical image-name pairs?), a paired-samples t -test was run to compare the percentage of easy and difficult items chosen for retrieval practice in the learning task. The results showed that participants chose

Table 1 Means (SDs) of delayed recall performance in Experiment 1 and Experiment 2

	Easy items	Difficult items
Experiment 1		
Restudy	53.66% (38.74%)	23.83% (27.34%)
Retrieval practice	79.94% (15.61%)	49.28% (32.88%)
Experiment 2		
Restudy	51.89% (36.23%)	20.90% (27.24%)
Retrieval practice	76.96% (22.98%)	34.69% (28.11%)

Note. As the learning strategy choice for each item was self-regulated, only parts of the participants chose restudy and retrieval practice for both easy and difficult items during the learning task and returned to complete the delayed test on Day 4. So, we reported the delayed recall performance based on these participants' data ($N = 28$ in Experiment 1, $N = 46$ in Experiment 2)

to take retrieval practice on a significantly higher percentage of easy items than difficult items, $t(69)=2.75$, $p=.01$, Cohen's $d=0.33$ (see Fig. 3).

As shown in Fig. 3, the percentage of easy and difficult items chosen for retrieval practice was heavily skewed. The Shapiro–Wilk test revealed that the distribution of retrieval practice choices for both easy and difficult items significantly departed from normality ($W_{\text{easy}}=0.78$, $W_{\text{difficult}}=0.81$, $ps<.001$). In this way, a non-parametric test was also conducted. A Wilcoxon signed-rank test further confirmed a significant difference between the percentage of easy items chosen for retrieval practice ($\text{Median}=90\%$) and the percentage of difficult items chosen for retrieval practice ($\text{Median}=82.5\%$), $Z=2.82$, $p=.005$.

The Influence of PME and JOL on Students' Learning Strategy Choices for Easy and Difficult Items

The Means (SDs) of participants' PME and JOL for easy items, difficult items and the average across all items are presented in Table 2. To address the second research question (i.e., How do PME and JOL associated with the items influence students' learning strategy choices for easy and difficult items?), we conducted a serial mediation model with item difficulty (0=easy item vs. 1=difficult item) as the predictor, PME as the first mediator, JOL as the second mediator, and learning strategy choice (0=restudy vs. 1=retrieval practice) as the binary outcome variable. Because the data have a multilevel structure (Level 1: items; Level 2: participants), we analyzed it with multilevel structural equation modeling (MSEM) in Mplus Version 8.8 Demo (Muthén & Muthén, 2017). MSEM is less prone to biases and has advantages over

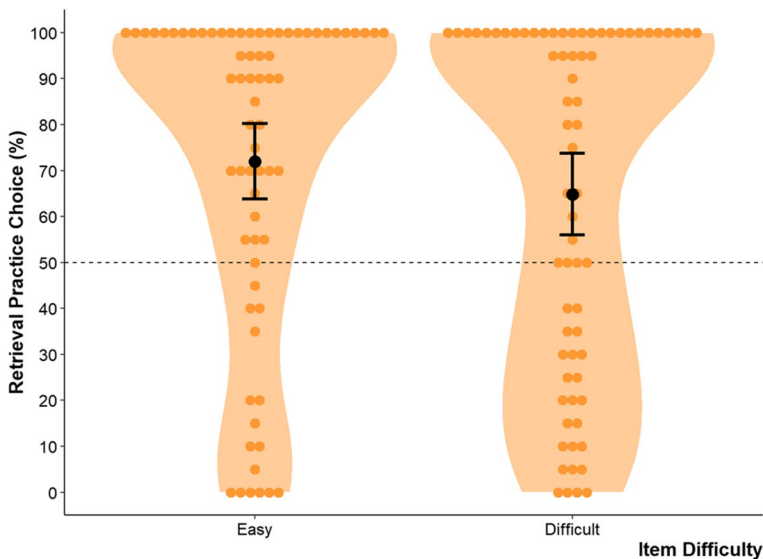


Fig. 3 Violin plot depicting the percentage of easy (and difficult) items chosen for retrieval practice among participants. Each orange dot represents one participant's data, and the black point represents the mean among participants. Error bars represent 95% confidence interval (CI)

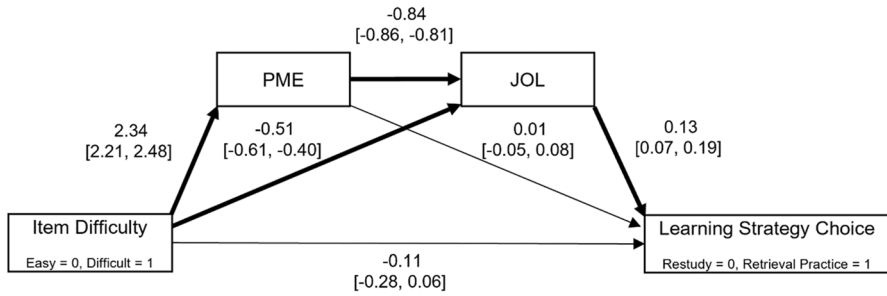
Table 2 Means (SDs) of PME and JOL in Experiments 1 and 2

	Easy items	Difficult items	Average
Experiment 1 (N = 70)			
PME	3.47 (1.48)	5.81 (1.84)	4.64 (1.56)
JOL	63.86 (13.78)	39.24 (15.46)	51.55 (13.22)
Experiment 2 (N = 120)			
The first learning task			
PME	3.70 (1.62)	6.04 (1.81)	4.87 (1.62)
JOL	61.47 (18.14)	38.65 (18.34)	50.06 (17.19)
The second learning task			
Control group (N=52)			
PME	3.96 (1.30)	6.21 (1.58)	5.09 (1.30)
JOL	58.46 (19.12)	36.09 (19.42)	47.27 (18.03)
Intervention group (N=53)			
PME	3.43 (1.50)	6.24 (1.64)	4.84 (1.43)
JOL	65.60 (16.11)	38.44 (16.88)	52.02 (14.70)

Note. PME = perceived mental effort; JOL = judgment of learning

multilevel modeling (MLM) for mediation in clustered data (Preacher et al., 2011). Based on Preacher et al.'s online syntax Model I, we conducted a 1–1–1 model with fixed slopes (MSEM). Learning strategy choice was a binary outcome, so we used Bayesian estimators with default (non-informative) priors and means for point estimates. Mplus uses the default of two Markov Chain Monte Carlo (MCMC) chains, and we used 100,000 iterations for each chain because the current model was complex (Muthén, 2010). JOL was transferred from 0 – 100% to 0 – 10% for a better model fit. Below, the item-level (i.e., within-level) results were reported because participants made learning strategy choices at item level.

As shown in Fig. 4, the current serial mediation model comprises three indirect effects and one direct effect for the effect of item difficulty on learning strategy choice. The total indirect effect was the sum of all three indirect effects, which was significant (-0.29 , 95% CI = $[-0.39, -0.20]$). Specifically, the indirect effect of item difficulty on learning strategy choice through PME was not significant (0.03 , 95% CI = $[-0.12, 0.18]$). This result may come from the non-significant influence of PME on learning strategy choice (0.01 , 95% CI = $[-0.05, 0.08]$). As shown in Fig. 4, the influence of PME on learning strategy choice was entirely mediated by JOL, resulting in a null direct effect of PME on learning strategy choice. The indirect effect of item difficulty on learning strategy choice through JOL was significant (-0.07 , 95% CI = $[-0.10, -0.04]$). Difficult items were rated with lower JOL compared to easy items, and in turn discouraged the choice of retrieval practice. Additionally, the effect of item difficulty on JOL was partially mediated by PME, leaving a significant direct effect of item difficulty on JOL (-0.51 , 95% CI = $[-0.61, -0.40]$). This resulted in a significant indirect effect of item difficulty on learning strategy choice through PME and JOL in series (-0.26 , 95% CI = $[-0.37, -0.14]$). When learning difficult items, participants perceived greater mental effort (compared to learning easy items), leading to lower JOL, and in the



Within-level indirect effects:

Item Difficulty → PME → Learning Strategy Choice: 0.03, 95% CI = [-0.12, 0.18]

Item Difficulty → JOL → Learning Strategy Choice: **-0.07, 95% CI = [-0.10, -0.04]**

Item Difficulty → PME → JOL → Learning Strategy Choice: **-0.26, 95% CI = [-0.37, -0.14]**

Total within level indirect effect: **-0.29, 95% CI = [-0.39, -0.20]**

Fig. 4 Multilevel serial mediation model showing the effect of item difficulty on learning strategy choice, as serially mediated by PME and JOL. Each estimate is presented with 95% credibility intervals in square brackets. The significant paths are shown in bold. Dummy coding for item difficulty and learning strategy choice are presented in the boxes

end decreased the likelihood of choosing retrieval practice. Finally, the direct effect of item difficulty on learning strategy choice was not significant when PME and JOL were controlled (-0.11 , 95% CI = $[-0.28, 0.06]$).

Learning Strategy Beliefs

Participants' learning strategy beliefs (i.e., effectiveness ratings) were analyzed using a 2 (strategy: restudy vs. retrieval practice) \times 3 (time: first vs. second vs. third) within-subjects repeated-measures ANOVA. The main effect of strategy was significant, $F(1, 62) = 79.07$, $p < .001$, $\eta_p^2 = 0.56$, indicating that participants overall rated retrieval practice as more effective than restudy. The main effect of time was significant, $F(2, 124) = 43.70$, $p < .001$, $\eta_p^2 = 0.41$, with effectiveness ratings decreasing across time (first > second > third). Furthermore, there was a significant interaction between strategy and time, $F(2, 124) = 11.73$, $p < .001$, $\eta_p^2 = 0.16$, suggesting that the changes in effectiveness ratings across time varied between retrieval practice and restudy (see Fig. 5a).

To illustrate this interaction, separate ANOVAs were conducted on the effectiveness ratings for retrieval practice and restudy, with time as the only within-subject variable. The main effect of time on the effectiveness ratings for retrieval practice was significant, $F(2, 124) = 6.88$, $p = .001$, $\eta_p^2 = 0.10$. Pairwise comparisons with Bonferroni adjustment revealed that participants' effectiveness ratings for retrieval practice remained stable after the learning task, $M_{\text{difference (first-second)}} = 0.10$, $SE_{\text{difference (first-second)}} = 0.15$, 95% CI $_{\text{difference (first-second)}} = [-0.26, 0.46]$, but decreased

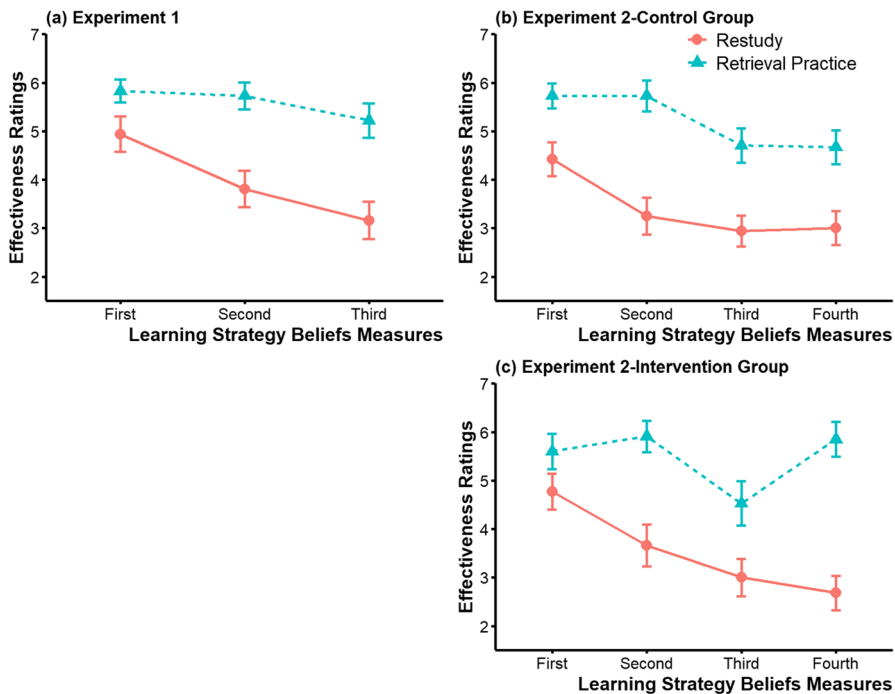


Fig. 5 Changes in the effectiveness ratings for restudy and retrieval practice strategies across time, in **a** Experiment 1 ($N=63$), **b** Experiment 2's control group ($N=52$), and **c** Experiment 2's intervention group ($N=53$). Error bars represent 95% CI. See the “Learning strategy beliefs measure” boxes in Fig. 1 to clarify when every measure occurs during the procedure

after the delayed test, $M_{\text{difference (second-third)}}=0.51$, $SE_{\text{difference (second-third)}}=0.17$, 95% CI $_{\text{difference (second-third)}}=[0.09, 0.93]$.

The main effect of time on the effectiveness ratings for restudy was significant, $F(2, 124)=43.87$, $p<.001$, $\eta_p^2=0.41$. Pairwise comparisons with Bonferroni adjustment revealed that participants' effectiveness ratings for restudy decreased after the learning task, $M_{\text{difference (first-second)}}=1.13$, $SE_{\text{difference (first-second)}}=0.19$, 95% CI $_{\text{difference (first-second)}}=[0.66, 1.60]$, and further decreased after the delayed test, $M_{\text{difference (second-third)}}=0.65$, $SE_{\text{difference (second-third)}}=0.19$, 95% CI $_{\text{difference (second-third)}}=[0.18, 1.12]$.

Discussion

Experiment 1 aimed to answer how students regulate their use of retrieval practice for easy and difficult anatomical image-name pairs and how PME and JOL associated with the items influence students' learning strategy choices for these items. First, Experiment 1 observed that participants chose to take retrieval practice less frequently for difficult items compared to easy ones, which confirmed Hypothesis 1. Additionally, Experiment 1 found that the influence of item

difficulty on learning strategy choice (retrieval practice *vs.* restudy) was mediated by (1) JOL alone and (2) PME and JOL in series, which confirmed Hypothesis 2. The task-level experiences (i.e., PME and JOL associated with the item) played central roles in guiding students' selective use of retrieval practice based on item difficulty. Participants experienced higher PME and lower JOL when learning difficult items, leading to a decreased inclination to choose retrieval practice (due to its effortfulness and lack of immediate learning gains).

Lastly, Experiment 1 found that although students generally perceived retrieval practice as more effective than restudying, they still employed retrieval practice *less* frequently for difficult items compared to easy ones. Rea et al. (2022) demonstrated that even though students recognize the effectiveness of retrieval practice, they may still resort to less effective strategies due to perceived time constraints and effort associated with effective learning strategies. It appears that students' awareness of effectiveness does not automatically translate into the adoption of effective learning strategies, especially when learning difficult study materials. Additionally, students decreased their effectiveness ratings for retrieval practice after the delayed test, suggesting a lack of full understanding regarding the effectiveness of retrieval practice in enhancing long-term retention. These findings collectively underscore the necessity for instructional interventions addressing the use of retrieval practice for items of varying difficulty.

Experiment 2

In Experiment 1, participants demonstrated a tendency to engage in retrieval practice less frequently for difficult items compared to easy ones. Despite extensive research in cognitive and educational psychology demonstrating retrieval practice as an effective learning strategy for enhancing long-term retention across varying difficulty levels (e.g., Rowland, 2014), students often underutilize this strategy, particularly for difficult materials. To address this, Experiment 2 aimed to investigate whether an instructional intervention on retrieval practice can improve the use of retrieval practice for both easy and difficult materials.

Participants in Experiment 2 were randomly assigned to one of the two groups: half received the instructional intervention (i.e., intervention group) while the other half received unrelated reading materials (i.e., control group). The intervention explicitly addressed two key points: (1) students selectively use learning strategy (retrieval practice *vs.* restudy) based on item difficulty, employing restudying more often for difficult items than for easy ones; (2) compared to restudying, retrieval practice benefits both easy and difficult items in long-term retention. In addition to retesting Hypotheses 1 and 2 from Experiment 1, we hypothesized that:

Hypothesis 3: Compared to the control group, participants in the intervention group increase the use of retrieval practice for both easy and difficult materials after the intervention.

Method

Participants

The sample size was determined based on the effect size of the difference in the percentage of pairs chosen for retrieval practice between easy and difficult items in Experiment 1 (Cohen's $d=0.33$), which suggests that 101 participants were required to achieve a power of 95% ($\alpha=0.05$). To be more conservative, we decided to increase the sample size to 120.

One hundred and twenty participants (85 females; age: $M=20.54$ years, $SD=1.36$) were recruited from Prolific for monetary compensation (£12.50; the 50% highest scoring participants on the memory test received an additional bonus of £1). All participants spoke English as a first language, did not major in medical or biological subjects, and did not report having reading disorders. Additionally, all participants had attained an educational level equal to or above high school graduate: high school graduate (24.2%), some college but no degree (54.2%), 2-year associate degree in college (7.5%), and 4-year Bachelor's degree in college (14.2%). 66.7% of participants had never learned human anatomy before, and the others had learned some before university. Since 92.5% of participants had never learned Latin while 97.5% of participants had never learned ancient Greek, the majority were not familiar with the anatomical names which are often closely related to Latin or ancient Greek. Participants were tested individually, and all provided online informed consent. All procedures were approved by the Ethics Review Committee of the Faculty of Health, Medicine, and Life sciences at Maastricht University (approval number: FHML-REC/2022/073), and informed consent was obtained from all participants.

Materials

In Experiment 2, alongside the 40 image-name pairs utilized in Experiment 1 (units 1–5), an additional forty new pairs were included for the second learning task (see Fig. 1 and Online Supplementary Materials: Appendix A), half of which were easy items, and the others were difficult ones. The item difficulty of both easy and difficult materials remained consistent with that of Experiment 1. The 40 anatomical image-name pairs were organized into five units (units 6–10), each comprising eight pairs (four easy and four difficult pairs).

Procedure

The procedure of Experiment 2 is shown in Fig. 1. On Day 1, the procedure was identical to that of Experiment 1. On Day 4, following participants' completion of the delayed recall test, their learning strategy beliefs were assessed for the third time (see the questions under the heading “learning strategy beliefs” in the “Measures” section of Experiment 1). Subsequently, half of the participants were randomly assigned to the intervention group. Another half of the participants were randomly assigned to the control group. Unlike the intervention group, the control group received a placebo on an unrelated reading text. The text was based on a report of PISA (Pál, 2018) and was

presented in a manner similar to that of the intervention group, including an equivalent number of figures (for the detailed instructions participants in both groups received, see Online Supplementary Materials: Appendix C). Participants were instructed to carefully read the intervention (or the placebo text). After that, participants' beliefs about retrieval practice and restudy were measured for the fourth time. Participants then engaged in the second learning task, during which they studied forty new image-name pairs (units 6–10). The learning process was identical to the first learning task (see Fig. 2).

After completing the second learning task, participants took a three-minute break to watch a cartoon. Afterward, participants took an immediate test on ten image-name pairs selected from the second learning task (five easy and five difficult pairs). This immediate test was designed to motivate participants to engage actively in the second learning task (e.g., students were informed that they would undergo an immediate memory test on the items studied in the second task), and its data will not be analyzed in the current study. At the end of Day 4, participants received feedback about their recall performance on both the delayed test and the immediate test, and they were asked to report any Internet connection problems, disturbances, or assistance from external sources during the test.

Results

All participants passed the attention checks throughout the experiment. Fifteen participants did not return for Day 4, resulting in a final sample size of 53 participants in the intervention group and 52 participants in the control group.

Delayed Recall Performance

Table 1 shows the *Means* (*SDs*) of the delayed recall performance. We conducted the same 2 (item difficulty: easy vs. difficult) \times 2 (strategy: restudy vs. retrieval practice) ANOVA as in Experiment 1 and replicated the results. The main effect of item difficulty was significant, $F(1, 45) = 123.66$, $p < .001$, $\eta_p^2 = 0.73$. Similarly, the main effect of strategy was significant, $F(1, 45) = 35.34$, $p < .001$, $\eta_p^2 = 0.44$. However, the item difficulty \times strategy interaction was not significant, $F(1, 45) = 2.62$, $p = .11$, $\eta_p^2 = 0.06$.

Retrieval Practice Choice

We summarized the *Means* (*SDs*) of the percentages of easy and difficult pairs chosen for retrieval practice in the first and second learning tasks for participants who did not return for Day 4, for those who were assigned to the control group on Day 4, and for those who were assigned to the intervention group on Day 4 (see Table 3).

Similar to Experiment 1, a paired-samples *t*-test was conducted to compare the percentage of easy and difficult items chosen for retrieval practice in the first learning task (i.e., before the intervention). The results showed that participants chose to take retrieval practice on numerically higher percentages of easy items ($M = 77.33\%$,

Table 3 Means (SDs) of the percentages of easy and difficult pairs chosen for retrieval practice in the first and second learning tasks in Experiment 2

	Easy items	Difficult items
Day 4 not back ($N=15$)		
First learning task	62.00% (41.35%)	59.67% (40.29%)
Intervention group ($N=53$)		
First learning task	77.64% (29.80%)	78.77% (30.44%)
Second learning task	86.32% (28.25%)	88.11% (26.37%)
Control group ($N=52$)		
First learning task	81.44% (29.08%)	77.40% (31.41%)
Second learning task	83.37% (30.18%)	78.46% (32.85%)

$SD=31.44\%$) compared to difficult items ($M=75.79\%$, $SD=32.51\%$). However, this difference was not significant, $t(119)=0.93$, $p=.35$, Cohen's $d=0.09$. Additionally, a Wilcoxon signed-rank test also revealed no significant difference between the percentage of easy items chosen for retrieval practice ($Median=95\%$) and the percentage of difficult items chosen for retrieval practice ($Median=95\%$), $Z=1.17$, $p=.24$. These results were discussed in detail in the “General Discussion” section.

To address the third research question (i.e., Does an instructional intervention on retrieval practice improve the use of retrieval practice for both easy and difficult materials?), participants' learning strategy choices for each item (0 = restudy, 1 = retrieval practice) were analyzed using a mixed-effects logistic regression analysis using the *lme4* and *lmerTest* packages in the R software (version 4.3.2). This analysis aimed to explicitly investigate whether participants in the intervention group, compared to those in the control group, were more likely to choose retrieval practice for both easy and difficult items in the second learning task after controlling for their choices in the first learning task. Fixed factors included item difficulty (easy vs. difficult), group (intervention vs. control), learning task phase (the first learning task vs. the second learning task), and their interactions. The random intercept was included in the model. The model and its outputs are presented in Table 4.

Two effects in Table 4 provided insight into the answer to our third research question. First, the effect of learning task phase (reference level = the first learning task) was significant ($\beta=1.12$, $OR=3.06$, $SE=0.15$, $z=7.28$, $p<.001$, 95% $CI=[0.82, 1.42]$). This indicated that from the first learning task to the second learning task, participants in the intervention group increased their odds of choosing retrieval practice for difficult items by 3.06 times.² Furthermore, the interaction of group and learning task phase was also significant ($\beta=-1.01$, $OR=0.37$, $SE=0.21$, $z=-4.85$, $p<.001$, 95% $CI=[-1.41, -0.60]$). This interaction suggested that the intervention group demonstrated a larger increase in the odds of choosing retrieval practice for difficult items from the first learning task (i.e., pre-intervention) to the second learning task (i.e.,

² When item difficulty was re-coded (reference level = easy) and the model was re-run, the effect of learning task phase remained significant ($\beta=0.97$, $OR=2.64$, $SE=0.15$, $z=6.61$, $p<.001$, 95% $CI=[0.68, 1.26]$). Now, this effect indicated that from the first learning task to the second learning task participants in the intervention group increased their odds of choosing retrieval practice for easy items by 2.64 times.

Table 4 Mixed-effects logistic regression analysis

Fixed effects	β	OR (e^{β})	SE	z value	p value	95% CI
(Intercept)	2.47	11.84	0.42	5.92	<.001	[1.65, 3.29]
Item difficulty (reference level = difficult)	-0.11	0.90	0.13	-0.83	.41	[-0.37, 0.15]
Group (reference level = intervention)	0.48	1.61	0.60	0.80	.42	[-0.69, 1.65]
Learning task phase (reference level = the first learning task)	1.12	3.06	0.15	7.28	<.001	[0.82, 1.42]
Item difficulty \times Group	0.55	1.73	0.19	2.80	.01	[0.16, 0.93]
Item difficulty \times Learning task phase	-0.15	0.86	0.21	-0.71	.48	[-0.56, 0.26]
Group \times Learning task phase	-1.01	0.37	0.21	-4.85	<.001	[-1.41, -0.60]
Item difficulty \times Group \times Learning task phase	0.26	1.30	0.29	0.89	.37	[-0.31, 0.83]

Note. OR = odds ratio; SE = standard error around the β estimates; 95% CI = 95% confidence interval around the β estimates. Model specification: glmer (Learning strategy choice ~ Item difficulty \times Group \times Learning task phase + (1 | participant), family = binomial (link = “logit”). For each effect, the reference level is indicated in the parentheses. Learning strategy choice: 0 = restudy, 1 = retrieval practice

post-intervention) compared to the control group.³ To aid readers in interpreting this interaction, we plotted the average percentages of easy and difficult items chosen for retrieval practice during the first learning task (pre-intervention) and the second learning task (post-intervention) in both the intervention group and the control group (see Fig. 6).

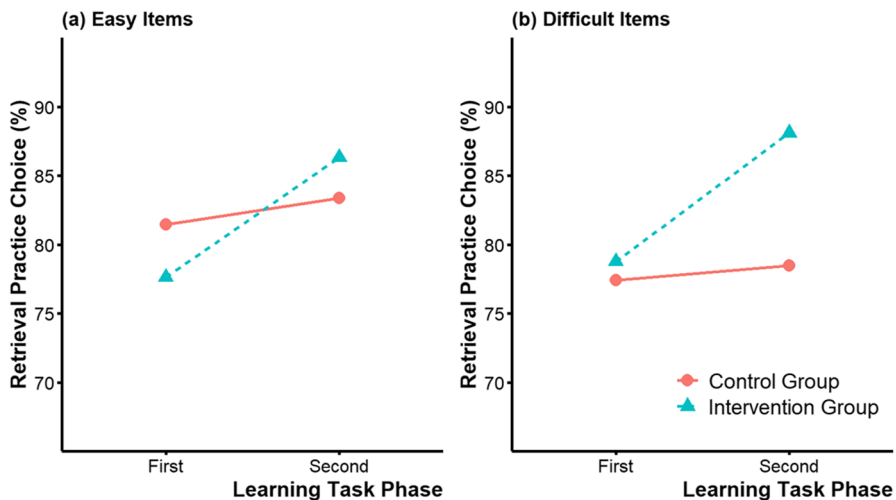


Fig. 6 The average percentages of **a** easy items and **b** difficult items chosen for retrieval practice during the first learning task (pre-intervention) and the second learning task (post-intervention) in the intervention group and the control group

³ When item difficulty was re-coded (reference level = easy) and the model was re-run, the interaction of *group* and *learning task phase* was also significant ($\beta = -0.75$, OR = 0.47, SE = 0.21, $z = -3.59$, $p < .001$, 95% CI = [-1.15, -0.34]). Now, this effect indicated that the intervention group demonstrated a larger increase in the odds of choosing retrieval practice for easy items from the first learning task to the second learning task compared to the control group.

Exploratory Analyses on the Influence of the Intervention on PME and JOL

Participants' mean ratings of PME and JOL for easy and difficult items during the two learning tasks are shown in Table 2. To explore the potential impact of the intervention on participants' ratings of PME and JOL during the second learning task (i.e., post-intervention), two 2 (item difficulty: easy vs. difficult) \times 2 (group: control vs. intervention) repeated-measures ANOVAs were conducted. Item difficulty served as a within-subjects variable, while group served as a between-subjects variable.

For participants' ratings of PME during the second learning task (i.e., post-intervention), the results showed that the main effect of item difficulty was significant, $F(1, 103) = 407.52$, $p < .001$, $\eta_p^2 = 0.80$, indicating that participants rated higher PME for difficult items than for easy ones during the second learning task. The main effect of group was not significant, $F(1, 103) = 0.88$, $p = .35$, $\eta_p^2 = 0.008$. The interaction between item difficulty and group was significant, $F(1, 103) = 4.97$, $p = .03$, $\eta_p^2 = 0.05$. This interaction revealed that PME for easy items were numerically lower in the intervention group compared to the control group, $t(103) = -1.93$, $p = .06$, Cohen's $d = -0.38$; however, PME for difficult items remained the same between the two groups, $t(103) = 0.10$, $p = .93$, Cohen's $d = 0.02$.

For participants' ratings of JOL during the second learning task (i.e., post-intervention), the results showed that the main effect of item difficulty was significant, $F(1, 103) = 314.43$, $p < .001$, $\eta_p^2 = 0.75$, indicating that participants rated lower JOL for difficult items than for easy ones during the second learning task. The main effect of group was not significant, $F(1, 103) = 2.19$, $p = .14$, $\eta_p^2 = 0.02$. Although the interaction between item difficulty and group was only marginal significant ($F(1, 103) = 2.94$, $p = .09$, $\eta_p^2 = 0.03$), JOL for easy items were higher in the intervention group compared to the control group, $t(103) = 2.07$, $p = .04$, Cohen's $d = 0.41$; however, JOL for difficult items remained the same between the two groups, $t(103) = 0.66$, $p = .51$, Cohen's $d = 0.13$.

To revisit the second research question (i.e., How do PME and JOL associated with the items influence students' learning strategy choices for easy and difficult items?), and further explore whether the intervention changes any relationships between variables (i.e., Item Difficulty \rightarrow PME \rightarrow JOL \rightarrow Learning Strategy Choice), the same multilevel serial mediation analyses were conducted as in Experiment 1 for both the first and the second learning tasks. The detailed results which may be of interest to readers are reported in the Online Supplementary Materials: see Appendix D. It seems that the intervention moderately changed the strength of several paths in the serial mediation model (e.g., Item Difficulty \rightarrow PME, PME \rightarrow Learning Strategy Choice). We further discussed these results in the [General Discussion](#).

Learning Strategy Beliefs

In Experiment 2, participants' learning strategy beliefs (i.e., effectiveness ratings) for retrieval practice and restudy were measured at four different time points. The

results from the first three time points replicated the same pattern as observed in Experiment 1 (see Fig. 5b, c). Experiment 2 then focused on the effectiveness ratings at the third measure (pre-intervention) and the fourth measure (post-intervention). A 2 (strategy: restudy vs. retrieval practice) \times 2 (time: third vs. fourth) \times 2 (group: control vs. intervention) mixed ANOVA was conducted, with strategy and time as the within-subjects variable and group as the between-subjects variable. The three-way interaction (i.e., strategy \times time \times group) was significant, $F(1, 103) = 19.68$, $p < .001$, $\eta_p^2 = 0.16$, indicating that the intervention group differed from the control group in the patterns of effectiveness ratings for restudy and retrieval practice from the third measure to the fourth measure. To ascertain the effectiveness of the intervention in enhancing participants' awareness of the effectiveness of retrieval practice, we performed two separate 2 (strategy: restudy vs. retrieval practice) \times 2 (time: third vs. fourth) within-subjects ANOVAs; one for the control group and one for the intervention group.

For the control group, the main effect of strategy was significant, $F(1, 51) = 66.74$, $p < .001$, $\eta_p^2 = 0.57$. However, neither the main effect of time nor the strategy \times time interaction was significant, indicating that the effectiveness ratings for both retrieval practice and restudy did not change between the third and fourth measures (see Fig. 5b). This result was anticipated because participants in the control group received instructions that were unrelated to learning strategy.

In contrast, for the intervention group, the main effect of strategy was significant, $F(1, 52) = 86.76$, $p < .001$, $\eta_p^2 = 0.63$. Moreover, the main effect of time was significant, $F(1, 52) = 17.90$, $p < .001$, $\eta_p^2 = 0.26$. Importantly, we observed a significant strategy \times time interaction, $F(1, 52) = 23.49$, $p < .001$, $\eta_p^2 = 0.31$. As shown in Fig. 5c, participants significantly increased their effectiveness ratings for retrieval practice after the intervention, $F(1, 52) = 33.03$, $p < .001$, which implied that the intervention effectively improved participants' awareness of the effectiveness of retrieval practice. However, participants only numerically decreased their effectiveness ratings for restudy after the intervention, $F(1, 52) = 3.17$, $p = .08$.

Discussion

Experiment 2 was designed to investigate whether an instructional intervention on retrieval practice can improve the use of retrieval practice for both easy and difficult materials. The results of Experiment 2 revealed two key findings: (1) participants in the intervention group increased their use of retrieval practice for both easy and difficult materials after the intervention, and (2) such increase in the odds of choosing retrieval practice for both types of items were more pronounced in the intervention group than in the control group. These results confirmed Hypothesis 3.

Without instruction, students often overlook the benefits of engaging with desirably difficult strategy and instead use the ineffective learning strategy, particularly when faced with challenging materials. The current study explicitly addressed such tendency of using more ineffective learning strategy (i.e., restudying) for difficult items, and then offered participants feedback about learning outcomes associated with retrieval practice and restudy for both easy and difficult materials, aiming to

educate them on the benefits of retrieval practice in fostering long-term retention for both types of materials. This instructional intervention was demonstrated to effectively enhance students' use of retrieval practice regardless of item difficulty.

General Discussion

Retrieval practice proves to be an effective learning strategy for enhancing students' long-term retention (Roediger & Butler, 2011). Previous studies have established that students selectively employ retrieval practice based on item difficulty, favoring restudying over retrieval practice for difficult learning materials (Badali et al., 2022; Toppino et al., 2018; Tullis et al., 2018). Building upon these studies, the present study aimed to provide new insights into how students regulate their use of retrieval practice when learning highly educationally relevant materials, such as human anatomical image-name pairs. Additionally, the current study explored how task-level experiences (i.e., PME and JOL associated with the items) influence students' learning strategy choices for easy and difficult items. Furthermore, the current study explored the potential efficacy of an instructional intervention in enhancing students' use of retrieval practice for materials of varying difficulty levels.

Students' Use of Retrieval Practice for Easy and Difficult Materials Before Instruction

Across two experiments, participants engaged in a four-round learning task. In the initial round, they studied both easy and difficult human anatomical image-name pairs and then decided whether to restudy or take retrieval practice on each item in the subsequent three rounds. The first main research question was: "How do students regulate their use of retrieval practice for easy and difficult materials?" Experiment 1 revealed that participants chose retrieval practice *less* frequently for difficult items compared to easy ones, thus confirming our hypothesis and aligning with prior studies (Badali et al., 2022; Toppino et al., 2018; Tullis et al., 2018). In Experiment 2, although students took retrieval practice on a numerically higher percentage of easy items than difficult ones, the difference was not significant.

In both experiments, even before any instructional intervention, participants consistently favored retrieval practice over restudying, with retrieval practice chosen over 60% of the time for both easy and difficult items (Experiment 1: $M_{\text{easy}} = 72\%$, $M_{\text{difficult}} = 64.86\%$; Experiment 2: $M_{\text{easy}} = 77.33\%$, $M_{\text{difficult}} = 75.79\%$). However, previous studies suggested that students tend to choose restudy over retrieval practice for difficult items. For example, participants initially chose retrieval practice only 48% of the time for difficult items in Badali et al.'s (2022) Experiment 1. The strong inclination toward retrieval practice by students in our study is encouraging concerning students' self-regulated learning behaviors. With the emergence of intervention studies aimed at enhancing students' use of effective learning strategies, there is a growing indication that students are becoming increasingly aware of the effectiveness of retrieval practice. Nevertheless, it is essential to be cautious in

interpreting the observed trend of selecting more retrieval practice for both types of materials in our study.

In the current study, participants received feedback about the correct name of each anatomical image during retrieval practice, enabling them to restudy materials and correct errors on incorrectly answered items. Providing feedback in retrieval practice allows participants to evaluate their learning by confirming correct answers and correcting errors on those incorrectly answered items. Therefore, retrieval practice *plus* feedback proves more beneficial than practice tests without feedback (Metcalfe et al., 2009). The provision of feedback may underscore the desirable aspects of retrieval practice, thereby motivating students to opt for retrieval practice. For example, Karpicke et al. (2009) conducted a survey on students' learning strategies, with one forced report question that asked students to choose whether they would restudy or take retrieval practice after studying a textbook chapter. Results showed that only 18% of participants chose retrieval practice in circumstance when not followed by a restudy opportunity (i.e., feedback), with the majority choosing restudy. However, when students could restudy after taking retrieval practice (i.e., retrieval practice with feedback), the proportion opting for retrieval practice increased from 18% to 42%. This study underscores the role of feedback in fostering students' inclination toward retrieval practice over restudy. Participants do not need to worry about forfeiting the chance to restudy if they cannot retrieve or retrieve an incorrect answer during retrieval practice because they always have the opportunity to correct themselves with immediate feedback about the correct answer. Therefore, providing feedback about the correct answer may lower the threshold of choosing retrieval practice and make retrieval practice more appealing than not providing feedback.

The Influence of PME and JOL on Students' Learning Strategy Choices for Easy and Difficult Items

The second main research question was: "How do PME and JOL associated with the items influence students' learning strategy choices for easy and difficult items?" Given that students selectively engage in *less* retrieval practice for difficult items, it is imperative to comprehend the underlying processes guiding such decisions.

Across two experiments, the results consistently showed that the influence of item difficulty on learning strategy choices (retrieval practice *vs.* restudy) was mediated by (1) JOL alone and (2) PME and JOL in series. Karpicke (2009) proposed that students' JOLs play a pivotal role in guiding strategy choices, thereby indicating the influence of metacognitive monitoring on metacognitive control (Nelson & Narens, 1994). In line with this perspective, the current study revealed that participants relied on their JOLs associated with the easy (or difficult) materials to inform their learning decisions (i.e., as indicated by "Item Difficulty → JOL → Learning Strategy Choice"). According to the cue-utilization framework, JOLs are inferential in nature and rely on various available cues. Item difficulty is an important cue that can influence JOL in a nonanalytic, experience-based way (Koriat, 1997). Difficult items were rated with lower JOLs. Participants may then weigh task-level experiences (e.g., JOL) against the strategy-level experiences (e.g., cognitive load

associated with retrieval practice or restudy strategy). As initial learning of difficult items demands considerable effort and occupies a significant portion of working memory capacity, with insufficient resources left, students may opt to avoid the effort-intensive retrieval practice strategy and instead utilize the restudy strategy (Seufert, 2018, 2020).

A plausible explanation for the serial mediating roles of PME and JOL in the relationship between item difficulty and learning strategy choice can be elucidated as follows. As shown in Fig. 4, the influence of item difficulty on JOL was also partially mediated by PME, in which students interpreted effort in a negative data-driven way (Baars et al., 2020; Kirk-Johnson et al., 2019; Koriat et al., 2014). Learning difficult items induced higher PME, leading to lower JOL; while learning easy items induced lower PME, leading to higher JOL. It is noteworthy that item difficulty did not moderate the relationship between PME and JOL, but instead served as a basic cue for both PME and JOL. Additionally, PME associated with the easy (or difficult) item could not directly guide students' decision-making regarding how to learn that item. Instead, the influence of PME on learning strategy choice was entirely mediated by JOL. Ultimately, students made learning strategy decisions by balancing between current task-level experiences (e.g., PME and JOL associated with the studied item) and the demands of the strategy itself. When task-level experiences incurred significant costs, such as experiencing greater PME and lower JOL when learning difficult items, students were less inclined to choose the retrieval practice strategy due to insufficient remaining resources. On the contrary, when task-level experiences did not impose significant costs on students, such as experiencing lower PME and higher JOL when learning easy items, students had sufficient remaining resources and were more likely to opt for the desirably difficult strategy—retrieval practice (Seufert, 2018, 2020).

Although it was not the primary focus of our intervention in Experiment 2, we further explored (1) the potential impact of the intervention on participants' ratings of PME and JOL during the second learning task (i.e., post-intervention) and (2) the potential impact of the intervention in the serial mediation model "Item Difficulty \rightarrow PME \rightarrow JOL \rightarrow Learning Strategy Choice." Two noteworthy changes were identified. Firstly, the intervention resulted in a numerical decrease in PME and an increase in JOL for easy items, but did not alter the ratings of PME and JOL for difficult items. It is possible that participants were already confident in learning easy items. After receiving an intervention that highlighted the benefits of retrieval practice for both easy and difficult items, they might believe that increased retrieval practice would further enhance their learning of easy items, thereby boosting their confidence in learning these items. This increased confidence in learning easy items may, in turn, influence their perceptions of how much effort was required to learn the easy items (PME) and how well they could learn them (JOL). However, this process may not apply to difficult items, as participants might find it harder to gain confidence in learning them. Even if participants believe that more retrieval practice is beneficial for learning difficult items, they might not significantly change their perceptions of the effort required to learn difficult items (PME) or how well they could learn them (JOL). Secondly, for the serial mediation model post-intervention, the

total within-level indirect effect, which is the sum of the three indirect effects (Item Difficulty \rightarrow PME \rightarrow Choice; Item Difficulty \rightarrow JOL \rightarrow Choice; Item Difficulty \rightarrow PME \rightarrow JOL \rightarrow Choice), was no longer significant and close to zero in the intervention group (-0.01), but was negative and significant in the control group (-0.36). Looking more closely into this effect, we found that the indirect effect “Item Difficulty \rightarrow PME \rightarrow Choice” drove this change. In the intervention group, this indirect effect was 0.33 , which counteracted the other two indirect effects (-0.04 and -0.30), resulting in the non-significant and close-to-zero total within-level indirect effect. Examining the indirect effect “Item Difficulty \rightarrow PME \rightarrow Choice” in detail, we observed that the effect of PME on learning strategy choice was positive and slightly larger than zero (although not significant) after the intervention (0.12 , 95% CI = $[-0.002, 0.24]$), but remained close to zero in the control group (-0.03 , 95% CI = $[-0.12, 0.06]$). After the intervention, there appears to be a weak inclination for PME to positively correlate with students’ learning strategy choice (i.e., retrieval practice vs. restudying). While students did not reduce their PME for difficult items following the intervention, they seemed to begin adopting the desirably difficult strategy (i.e., retrieval practice) instead of restudying. Although our intervention appears to decrease PME and increase JOL for easy items, as well as alter the relationship between PME and learning strategy choice, we should be cautious in interpreting these findings, and attempt to replicate them first in a future study.

Nevertheless, the current intervention did not directly address students’ on-task experiences, but doing so could be promising to further increase use of retrieval practice. Recently, Onan et al. (2024) developed an intervention to increase the use of the desirably difficult strategy—interleaved practice, which combines a theory-based method (e.g., refutations) and experience-based methods (e.g., strategy implementation and visual metacognitive prompts). Participants monitored their on-task experiences (i.e., perceived effort and learning) with blocked and interleaved practice and the visual metacognitive prompts directly provided participants opportunities to reflect on the changes of their on-task experiences with both strategies, through which participants could recognize the long-term benefits of interleaved practice and verify the content of refutations. Their study showed that the combination of refutations and metacognitive prompts form a strong intervention and increase students’ use of interleaved practice. Onan et al.’s study shows how an intervention can be designed to directly target on-task experiences and their relation to learning strategy choice. Future studies are welcomed to test whether similar interventions on retrieval practice could affect the relationships between on-task experiences (PME and JOL) and promote the use of retrieval practice. For example, whether students could adopt a positive attitude toward investing effort and stick to desirably difficult strategy (i.e., retrieval practice) for challenging materials after such intervention (de Bruin et al., 2023).

The Effect of an Instructional Intervention on Promoting the Use of Retrieval Practice

Our third main research question aimed to investigate: “Does an instructional intervention improve the use of retrieval practice for both easy and difficult materials?” The current study implemented an intervention wherein students received instructions indicating that (1) students selectively use learning strategy based on item difficulty, using more restudying for difficult items compared to easy ones, and (2) compared with restudying, retrieval practice benefits both easy and difficult items in long-term retention. Our findings demonstrated that this intervention led to an enhancement in the utilization of retrieval practice in a subsequent learning task. Specifically, there was a notable increase in the odds of choosing retrieval practice for both easy and difficult materials after the intervention. Moreover, the increase in the odds of choosing retrieval practice for both easy and difficult items was larger in the intervention group compared to the control group. These results suggested that an instructional intervention can effectively facilitate the self-adoption of the retrieval practice strategy.

Furthermore, although the current study found that students initially perceived retrieval practice as more effective than restudy, as shown in Fig. 5, this perception alone may not suffice to ensure its consistent use for items with varying difficulty. Additionally, students decreased their effectiveness ratings for retrieval practice after the delayed test, indicating that they may not fully understand the effectiveness of retrieval practice in boosting long-term retention. Importantly, the instructional intervention significantly boosted participants’ ratings of retrieval practice effectiveness post-intervention, indicating its efficacy in raising awareness toward the desirably difficult strategy.

However, the current study was conducted in specific experimental settings rather than in real educational environments, so its implications for educational practice should be further explored. Participants in experimental settings may behave differently than they would in actual educational contexts (Mitchell, 2012). For example, the current set of Prolific participants might be more inclined to follow instructions and engage in retrieval practice. However, in a real classroom, simply informing learners about the benefits of retrieval practice for both easy and difficult items will likely not lead to consistent use (Carpenter, 2023). McDaniel and Einstein (2020) recently proposed the Knowledge, Belief, Commitment, and Planning (KBCP) framework, which emphasizes the importance of implementing these four components together to guide the training of effective learning strategies. Additionally, the Start and Stick to Desirable Difficulties (S2D2) framework (de Bruin et al., 2023) highlights that beyond acquiring accurate knowledge of desirable learning strategies, effectively managing the perceived effort associated with these strategies through repeated practice with the strategies is crucial for self-regulated learning. We encourage future studies to investigate how to support students in increasing their use of retrieval practice for content of varying difficulty in real classroom settings, combining the current intervention with the key components proposed by the aforementioned theoretical frameworks.

Limitations and Future Studies

Several limitations should be considered in the current study. First, the experimental design organized anatomical image-name pairs into five units, each containing eight pairs. It is plausible that the size of the unit (i.e., the number of pairs within the unit) could influence students' learning strategy decisions, with fewer pairs potentially encouraging retrieval practice attempts rather than immediate access to correct answers. Future research should explore how the presentation format of organizing learning materials impacts students' learning strategy choices in self-regulated learning.

Second, the current study only investigated the short-term benefits of the instructional intervention, as participants were asked to engage in a new learning task immediately after the intervention. Investigating the long-lasting effectiveness of the instructional intervention in the future is worthwhile. Furthermore, although the current study showed that the instructional intervention can increase the use of retrieval practice for both easy and difficult materials, we did not assess whether the intervention had a positive impact on long-term retention for both types of materials. We advocate for future research to address this significant aspect.

Lastly, although the current study collected task-level experiences (e.g., PME and JOL associated with easy and difficult items), we did not collect strategy-level experiences at the meantime. For example, we did not gather data on students' perceived effort and learning associated with retrieval practice or restudying strategy, as previous studies did (Hui et al., 2022; Kirk-Johnson et al., 2019). Hence, the current study lacks direct empirical evidence for the argument that students rely on a balance between task-level experiences and strategy-level experiences when deciding whether to restudy or engage in retrieval practice for each easy or difficult item. Future study is encouraged to empirically test such argument.

Conclusion

The present study found that students selectively employ retrieval practice based on item difficulty, using it less frequently for difficult items compared to easy ones. Moreover, students' task-level experiences such as PME and JOL influence their learning strategy choices for easy and difficult items. More essentially, the current study was the first to find that an instructional intervention, which revealed that while students prefer restudying for complex items, retrieval practice benefits both easy and difficult items in long-term retention, could improve students' use of retrieval practice for both types of materials. In conclusion, the instructional intervention could serve as an effective tool to promote students' self-adoptions of retrieval practice regardless of item difficulty.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-024-09945-3>.

Funding This research was funded by the Dutch Research Council (NWO) (VIDI grant number: VI.Vidi.195.135), the National Natural Science Foundation of China (grant number: 32171045), the Fundamental Research Funds for the Central Universities (grant number: 2243300005) and the China Scholarship Council (File No. 202006040180).

Data Availability The data for the current study are available from the corresponding author upon request.

Declarations

Conflict of Interest The authors declare no competing interests.

References

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. <https://doi.org/10.3102/0034654316689306>
- Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied*, 24(1), 43–56. <https://doi.org/10.1037/xap0000133>
- Baars, M., Wijnia, L., de Bruin, A. B. H., & Paas, F. (2020). The relation between student's effort and monitoring judgments during learning: A meta-analysis. *Educational Psychology Review*, 1–24. <https://doi.org/10.1007/s10648-020-09569-3>
- Badali, S., Rawson, K. A., & Dunlosky, J. (2022). Do students effectively regulate their use of self-testing as a function of item difficulty? *Educational Psychology Review*, 34(3), 1651–1677. <https://doi.org/10.1007/s10648-022-09665-6>
- Biwer, F., oudeEgbrink, M. G. A., Aalten, P., & de Bruin, A. B. H. (2020). Fostering effective learning strategies in higher education—A mixed-methods study. *Journal of Applied Research in Memory and Cognition*, 9(2), 186–203. <https://doi.org/10.1016/j.jarmac.2020.03.004>
- Biwer, F., de Bruin, A. B. H., & Persky, A. (2022). Study smart – Impact of a learning strategy training on students' study behavior and academic performance. *Advances in Health Sciences Education*, 28(1), 147–167. <https://doi.org/10.1007/s10459-022-10149-z>
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444. <https://doi.org/10.1146/annurev-psych-113011-143823>
- Broeren, M., Heijltjes, A., Verkoeijen, P., Smeets, G., & Arends, L. (2021). Supporting the self-regulated use of retrieval practice: A higher education classroom experiment. *Contemporary Educational Psychology*, 64, 101939. <https://doi.org/10.1016/j.cedpsych.2020.101939>
- 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. <https://doi.org/10.1037/a0017021>
- Carpenter, S. K. (2023). Encouraging students to use retrieval practice: A review of emerging research from five types of interventions. *Educational Psychology Review*, 35(4), 96. <https://doi.org/10.1007/s10648-023-09811-8>
- Carpenter, S. K., Rahman, S., Lund, T. J. S., Armstrong, P. I., Lamm, M. H., Reason, R. D., & Coffman, C. R. (2017). Students' use of optional online reviews and its relationship to summative assessment outcomes in introductory biology. *CBE Life Sciences Education*, 16(2), ar23. <https://doi.org/10.1187/cbe.16-06-0205>
- Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, 1–16. <https://doi.org/10.1038/s44159-022-00089-1>

- de Jonge, M., & Tabbers, H. K. (2013). Repeated testing, item selection, and relearning: The benefits of testing outweigh the costs. *Experimental Psychology*, 60(3), 206–212. <https://doi.org/10.1027/1618-3169/a000189>
- de Bruin, A. B. H., Biwer, F., Hui, L., Onan, E., David, L., & Wiradhandy, W. (2023). Worth the effort: The start and stick to desirable difficulties (S2D2) framework. *Educational Psychology Review*, 35(2), 41. <https://doi.org/10.1007/s10648-023-09766-w>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, 19(1), 126–134. <https://doi.org/10.3758/s13423-011-0181-y>
- Hui, L., de Bruin, A. B. H., Donkers, J., & van Merriënboer, J. J. G. (2021). Does individual performance feedback increase the use of retrieval practice? *Educational Psychology Review*, 1, 1–23. <https://doi.org/10.1007/s10648-021-09604-x>
- Hui, L., de Bruin, A. B. H., Donkers, J., & van Merriënboer, J. J. G. (2022). Why students do (or do not) choose retrieval practice: Their perceptions of mental effort during task performance matter. *Applied Cognitive Psychology*, 36(2), 433–444. <https://doi.org/10.1002/acp.3933>
- Jang, Y., Wixted, J., Pecher, D., Zeelenberg, R., & Huber, D. (2011). Decomposing the interaction between retention interval and study/test practice: The role of retrievability. *Quarterly Journal of Experimental Psychology*, 65, 962–975. <https://doi.org/10.1080/17470218.2011.638079>
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138(4), 469. <https://doi.org/10.1037/a0017341>
- Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17(4), 471–479. <https://doi.org/10.1080/09658210802647009>
- 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. <https://doi.org/10.1016/j.cogpsych.2019.101237>
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349. <https://doi.org/10.1037/0096-3445.126.4.349>
- Koriat, A., Nussinson, R., & Ackerman, R. (2014). Judgments of learning depend on how learners interpret study effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1624–1637. <https://doi.org/10.1037/xlm0000009>
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, 14(2), 219–224. <https://doi.org/10.3758/BF03194055>
- McDaniel, M. A., & Einstein, G. O. (2020). Training learning strategies to promote self-regulation and transfer: The knowledge, belief, commitment, and planning framework. *Perspectives on Psychological Science*, 15(6), 1363–1381. <https://doi.org/10.1177/1745691620920723>
- Metcalf, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, 37(8), 1077–1087. <https://doi.org/10.3758/MC.37.8.1077>
- Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(9), 1474–1486. <https://doi.org/10.1037/xlm0000486>
- Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science*, 7(2), 109–117. <https://doi.org/10.1177/1745691611432343>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Author.
- Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction* (Technical Report). Retrieved from <https://www.statmodel.com/download/IntroBayesVersion%203.pdf>
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. *Metacognition: Knowing About Knowing*, 13, 1–25. <https://doi.org/10.7551/mitpress/4561.003.0003>

- Onan, E., Biwer, F., Abel, R., Wiradhany, W., & de Bruin, A. (2024). Optimizing self-organized study orders: Combining refutations and metacognitive prompts improves the use of interleaved practice. *NPIJ Science of Learning*, 9(1), 33. <https://doi.org/10.1038/s41539-024-00245-7>
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>
- Pál, J. (2018). How is participation in sports related to students' performance and well-being? *PISA in Focus*, No. 86. OECD Publishing, Paris. <https://doi.org/10.1787/e124db26-en>
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, 18(2), 161–182. <https://doi.org/10.1080/10705511.2011.557329>
- Rea, S. D., Wang, L., Muenks, K., & Yan, V. X. (2022). Students can (mostly) recognize effective learning, So why do they not do it?. *Journal of Intelligence*, 10(4), 127. <https://www.mdpi.com/2079-3200/10/4/127>
- Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review*, 33(3), 823–862. <https://doi.org/10.1007/s10648-020-09578-2>
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Roediger, 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. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. <https://doi.org/10.1037/a0037559>
- Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educational Research Review*, 24, 116–129. <https://doi.org/10.1016/j.edurev.2018.03.004>
- Seufert, T. (2020). Building bridges between self-regulation and cognitive load—An invitation for a broad and differentiated attempt. *Educational Psychology Review*, 32(4), 1151–1162. <https://doi.org/10.1007/s10648-020-09574-6>
- Toppino, T. C., LaVan, M. H., & Iaconelli, R. T. (2018). Metacognitive control in self-regulated learning: Conditions affecting the choice of restudying versus retrieval practice. *Memory & Cognition*, 46(7), 1164–1177. <https://doi.org/10.3758/s13421-018-0828-2>
- Tullis, J. G., Fiechter, J. L., & Benjamin, A. S. (2018). The efficacy of learners' testing choices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 540–552. <https://doi.org/10.1037/xlm0000473>
- Tullis, J. G., & Maddox, G. B. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition and Learning*, 15, 129–154. <https://doi.org/10.1007/s11409-020-09223-x>
- van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, 43(1), 16–26. <https://doi.org/10.1080/00461520701756248>
- Yan, V. X., Thai, K.-P., & Bjork, R. A. (2014). Habits and beliefs that guide self-regulated learning: Do they vary with mindset? *Journal of Applied Research in Memory and Cognition*, 3(3), 140–152. <https://doi.org/10.1016/j.jarmac.2014.04.003>
- Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, 147(4), 399–435. <https://doi.org/10.1037/bul0000309>
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995–1008. <https://doi.org/10.3758/MC.38.8.995>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Tian Fan¹  · Luotong Hui^{2,3}  · Liang Luo⁴  · Anique B. H. de Bruin² 

✉ Tian Fan
fantian@mail.bnu.edu.cn

¹ Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, No. 19 Xijiekouwai Street, Haidian District, Beijing 100875, China

² Department of Educational Development and Research, School of Health Professions Education, Maastricht University, Maastricht, The Netherlands

³ Imperial College London, London, UK

⁴ Institute of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China