



Evaluating the Judgment of Learning: its Limited Impact and the Power of Retrieval on Inductive Learning

Hyorim Ha¹ · Hee Seung Lee¹

Accepted: 21 May 2024 / Published online: 7 June 2024

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

Abstract

Recent studies suggest that making judgments of learning (JOLs)—self-assessment of current learning status—may not merely be a neutral cognitive process, but can directly improve learning through what is called ‘JOL reactivity’. This study investigated whether making JOLs can facilitate the learning of previously studied materials (backward effect) and newly studied materials (forward effect) in inductive learning. We also examined how this effect varies depending on whether a JOL is accompanied by a retrieval attempt. Across three experiments, participants learned about various butterfly species presented in two sections (Sections A and B). Some participants made JOLs between Section A and Section B, while others did not, and then all participants took a final transfer test for both sections. In Experiment 1, merely making JOLs did not facilitate learning compared to restudy control, regardless of whether JOLs afforded covert retrieval (target-absent JOL) or not (target-present JOL). However, in Experiment 2, when participants made JOLs combined with overt retrieval prompts (retrieval practice + JOL), they outperformed the other groups in the final transfer test of Section B, showing the forward effect. Experiment 3 further revealed that the act of making JOLs combined with overt retrieval practice was as effective as (but not more than) retrieval practice without JOLs in promoting new learning. Our findings indicate that conventional forms of JOLs do not appear to enhance inductive learning; rather, they underscore the critical role of retrieval in facilitating inductive learning.

Keywords Judgments of learning · Metacognition · JOL reactivity · Retrieval · Inductive learning

Metacognition is crucial to successful learning. As students evaluate their learning progress, they allocate their study time (Ariel et al., 2009; Son & Metcalfe, 2000),

✉ Hee Seung Lee
hslee00@yonsei.ac.kr

¹ Department of Education, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Korea

decide whether to continue or stop studying (Metcalfe & Finn, 2008), and/or change their study strategies (Sahakyan et al., 2004). Metacognitive judgment is often assumed to indirectly affect learning by influencing learners' study behavior (Nelson & Narens, 1990; Winne & Hadwin, 1998). However, recent studies suggest that the simple act of making a judgment of learning (JOL)—self-assessment of learning status—can also directly influence learning outcomes (e.g., Fiacconi et al., 2020; Janes et al., 2018; Kubik et al., 2022; Li et al., 2023; Maxwell & Huff, 2022; Soderstrom et al., 2015; Witherby et al., 2023; Zhao et al., 2022). For example, Soderstrom et al. (2015) had participants learn pairs of related words (e.g., herbs and spices) and unrelated words (e.g., nets and plants), exhibiting that making JOLs facilitated cued-recall performance, especially when the word pairs were strongly related, suggesting that a JOL itself can work as a memory modifier. This positive JOL reactivity has also been observed in other learning materials, including word pairs (Dougherty et al., 2005), word lists (Kubik et al., 2022; Li et al., 2021; Zhao et al., 2022), and physiological information (Dobson et al., 2019). These studies suggest that making JOLs is not a neutral cognitive process, but rather can directly affect learning through what is called 'JOL reactivity' (for reviews, see Double et al., 2018; Double & Birney, 2019).

Although an emerging number of studies have demonstrated the direct effects of JOLs on learning, one of the limitations of prior research is that most previous studies reporting positive reactivity examined the effect of JOL on learning that requires the retention of simple materials. Indeed, the meta-analysis conducted by Double et al. (2018) showed that the effect of making JOLs was positive, particularly for the retention of word lists ($g = 0.384$, 95% CI 0.146 to 0.622) and related word pairs ($g = 0.323$, 95% CI 0.083 to 0.563), while the overall effect of providing JOLs was not significant. This suggests that JOL reactivity may change depending on the type of learning material, especially when learning requires a higher level of knowledge acquisition beyond remembering simple facts. In educational contexts, most learning involves complex materials that often require understanding, generalization, and memorization. However, only a few studies have directly tested JOL reactivity using educationally relevant materials (e.g., Ariel et al., 2021; Dobson et al., 2019; Zhao et al., 2023). For example, Dobson et al. (2019) had participants learn a series of physiological texts using various strategies and revealed that a JOL task was more beneficial for the transfer of physiological information than a study-only strategy. Recently, Ariel et al. (2021) and Zhao et al. (2023) examined the effect of making JOLs on text comprehension. However, in these studies, the criterion test included only factual questions that could be answered based on recall of the information.

In the present study, thus, we aimed to investigate whether JOL reactivity occurs with inductive learning, which involves extracting patterns from studied materials and applying the acquired knowledge to other instances (Holland et al., 1989). Specifically, we focused on natural category learning. Category learning plays an essential role in accumulating and expanding knowledge by helping learners transfer what they have studied to new situations (Murphy, 2002). However, in educational settings, many learners have difficulty studying categories of natural objects (e.g., butterflies and rocks) as they often have nebulous and unclear characteristics, which can ultimately make clear distinctions challenging (Miyatsu et al., 2019). In the field

of category learning, previous studies have examined the effects of metacognitive judgments, mostly in the form of category learning judgment (CLJ). Making CLJs is known to have an indirect effect on aspects of learning, such as restudy selection (Hartwig & Dunlosky, 2017; Morehead et al., 2016). For instance, when students prepare for a biology test, they can make CLJs to decide which category is mastered (e.g., eastern tiger butterfly) or not (e.g., painted lady butterfly) to adjust their subsequent study time and strategies. However, these studies mainly investigated the indirect effect of metacognitive judgments on learning (e.g., selecting subsequent learning strategies) and did not directly investigate the effect of making a JOL itself on learning outcomes. Thus, the current study aimed to determine whether making JOLs can facilitate inductive learning by using natural butterfly species and to explore JOL as a potential study strategy to enhance inductive learning.

In addition to the type of learning material, another factor that may affect JOL reactivity is the likelihood that a JOL task prompts the retrieval process in learners. In the standard form of JOL, learners study complete information and evaluate the likelihood that they will remember the information learned in future examinations. Here, the availability of retrieval opportunities may depend on whether complete information is present or absent while making JOLs. For example, after learning a series of cue-target pairs (e.g., tree-lawn), learners are asked to judge the likelihood of their ability to retrieve the target information. If both cues and targets are present (e.g., tree-lawn) while making a JOL, learners are unlikely to retrieve the target. Conversely, if a target is absent (e.g., tree-?), learners are more likely to engage in the retrieval process and make JOLs based on the retrievability of the material.

The *covert retrieval hypothesis* suggests that the act of making JOLs allows learners to engage in a covert retrieval process, calling information to mind, or rehearsing previously learned information (Jönsson et al., 2012; Kimball & Metcalfe, 2003; Son & Metcalfe, 2005; Senkova & Otani, 2021). Learners strengthen their memory trace during the retrieval process, thereby enhancing their learning performance. For example, Dougherty et al. (2005) demonstrated that the covert retrieval process triggered by the act of making JOLs led to better learning performance. Similarly, Soderstrom et al. (2015) reported positive JOL reactivity when learners were asked to make target-absent JOLs for related word pairs, suggesting that performing target-absent JOLs likely evokes a covert retrieval process. Also, during the covert retrieval process, learners may experience retrieval failure, leading them to invest greater effort in subsequent learning. For example, Kubik et al. (2022) demonstrated that making JOLs with incomplete information, as opposed to complete information, facilitated new verbal learning compared to restudying, indicating that JOL reactivity depends on the extent to which it elicits covert retrieval attempts.

However, if learners do not attempt the retrieval process, making JOLs may not facilitate learning, which does not result in positive reactivity. The *truncated search hypothesis* suggests that learners rarely attempt retrieval practice while making JOLs (Tauber et al., 2015) and that such a lack of retrieval attempt does not enhance learning. According to the two-stage process theory (Son & Metcalfe, 2005), making JOLs does not necessarily evoke retrieval attempts. Learners often truncate the retrieval process, making metacognitive judgments using intuitive cues such as familiarity (Son & Metcalfe, 2005) and ease of access (Tauber et al., 2015).

Witherby and Tauber (2017) also noted that learners might not put as much effort into making JOLs as practicing retrieval (e.g., self-testing). Indeed, several recent studies have demonstrated that students do not always engage in “a full-blown covert retrieval attempt” (Tekin & Roediger, 2021, p. 407) even when the form of JOLs is framed to afford covert retrieval (Ariel et al., 2021; Tekin & Roediger, 2021). For instance, Ariel et al. (2021) reported that a standard form of JOLs without explicit retrieval practice did not produce learning gains in text comprehension regardless of whether it afforded covert retrieval (target-absent JOL) or not (target-present JOL). Only JOLs followed by overt retrieval practice produced learning benefits, suggesting that if learners do not spontaneously attempt retrieval, making JOLs is unlikely to facilitate learning. Accordingly, the current study investigated whether retrieval plays a pivotal role in JOL reactivity by varying the types of JOLs to afford covert or overt retrieval practice across the experiments.

Making JOLs can positively affect learning in two ways. On one hand, making JOLs can enhance the learning of previously studied materials, which is known as the *backward effect* (Ha & Lee, 2023; Lee & Ha, 2019; see also Pastötter & Bäuml, 2014 for backward testing effect). Most previous studies on JOL reactivity have focused on the backward effect of JOLs (e.g., Janes et al., 2018; Soderstrom et al., 2015; Tauber & Witherby, 2019; Witherby et al., 2023) by demonstrating the improved retention of the judged materials. On the other hand, JOL reactivity can occur by facilitating new learning, known as the *forward effect* (Kubik et al., 2022; Ha & Lee, 2023; Lee & Ha, 2019; see also Pastötter & Bäuml, 2014 for forward testing effect). Lee and Ha (2019) were the first to investigate the forward effect of metacognitive judgments. In their study, the participants who made JOLs on previously studied painting styles outperformed the restudy control group in learning new painting styles. More recently, Ha and Lee (2023) reported the forward effect of JOLs on text learning. However, these two studies did not directly examine the effect of JOL itself, but rather the effect of guided metacognitive judgments that involved three phases of judgments: monitoring, evaluating, and planning. Furthermore, these studies did not systematically examine the role of retrieval, leaving unanswered questions about why the forward effects of metacognitive judgments occurred. Therefore, the current study investigates both the backward and forward effects of making JOLs on inductive learning while controlling for retrieval opportunities.

Because no research has examined the direct consequences of making JOLs on inductive category learning, we lack a specific hypothesis regarding positive or negative JOL reactivity. On the one hand, we expect that the act of making JOLs will benefit inductive learning, as evidenced by several studies reporting positive JOL reactivity in verbal learning (Janes et al., 2018; Soderstrom et al., 2015). Moreover, given that numerous studies have demonstrated indirect benefits of metacognitive judgments on learning, making JOLs could positively affect learners' study regulation (Sahakyan et al., 2004; Yang et al., 2017), leading to better learning outcomes. On the other hand, we also have plausible reasons to expect that JOLs might not necessarily benefit inductive learning. The JOL reactivity may vary depending on the type of learning materials involved. In simple verbal learning tasks, such as recalling word pairs, learners are required to memorize the given learning materials, whereas in complex learning tasks, such as classifying natural categories, learners must first

identify category characteristics before memorizing them (Murphy, 2002). Thus, the act of making JOLs might not suffice to enhance learning that requires a higher level of knowledge acquisition, such as categorization.

Despite our uncertainty regarding the impact of making JOLs on inductive learning, we anticipate that if it does indeed facilitate inductive learning, the benefits will be more pronounced when JOL affords the retrieval process. A large number of studies have shown the power of retrieval on various types of learning, from simple verbal learning to high-level inductive learning tasks (e.g., Carpenter & DeLosh, 2006; Jacoby et al., 2010). Besides, several studies reported the benefits of retrieval increased with greater retrieval effort (Carpenter & DeLosh, 2006; Kang et al., 2007). Therefore, we expect that the stronger the retrieval is afforded by JOLs, the better the learning outcomes will be in inductive learning.

The Present Study

This study had two major goals. The first goal was to test in a straightforward manner whether making JOLs regarding previously learned materials would improve inductive learning. Specifically, we examined how the act of making JOLs affected the transfer of previously learned categories (backward effect) and subsequently studied new ones (forward effect) by adopting a procedure typically used in studies on the forward testing effect (Choi & Lee, 2020; Kang et al., 2023; Lee & Ahn, 2018; Lee & Ha, 2019). The second goal was to investigate the role of retrieval in JOL reactivity. We examined how the effect of making JOLs varied in inductive learning depending on the forms of JOLs that were framed to afford a covert or overt retrieval opportunity.

Figure 1 represents an overview of the procedures used in Experiments 1 to 3. Participants were instructed to study images of various butterfly species across two sections: six species each in Section A and Section B. After studying Section A, participants were instructed to perform one of the different interim activities: they made JOLs or did not (i.e., interim-restudy or retrieval practice) for the learned butterfly images of Section A. On completion of the interim activity, all participants studied new species of butterflies in Section B. After studying Section B, participants took a final transfer test in which they had to classify new examples of butterfly species learned from Sections A and B. Hence, the final test required not only remembering characteristics of butterfly species but also the generalization of the learned knowledge. Participants would only be able to classify new butterflies from the learned category successfully if they had studied the butterflies at a category level. If a backward effect of JOL occurred, the JOL group would outperform the restudy control group in the final test of Section A. If the forward effect of JOL occurred, the JOL group would outperform the control group in the final test of Section B.

Across the three experiments, we varied the forms of JOLs to investigate the role of covert or overt retrieval attempts. In Experiment 1, we tested the effects of two different forms of JOLs—target-absent and target-present. We compared these forms to a restudy control to determine whether JOLs framed to encourage covert retrieval enhance inductive learning. In the target-absent JOL condition, the category name

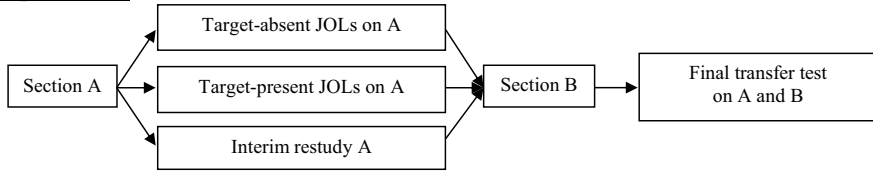
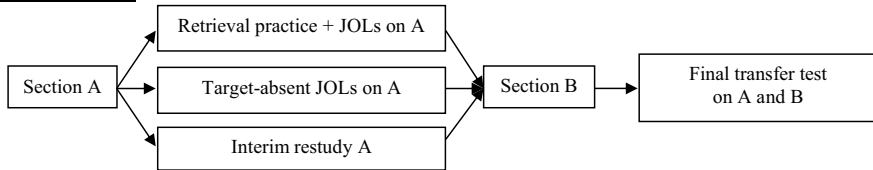
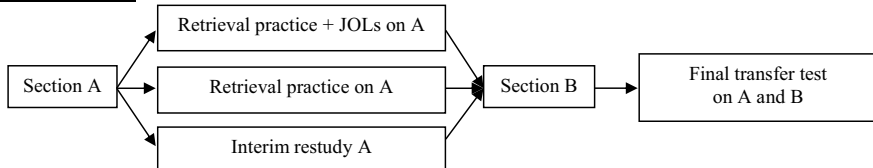
Experiment 1**Experiment 2****Experiment 3**

Fig. 1 An overview of procedures used in Experiments 1–3

of the butterfly was not provided; thus, participants were likely to engage in covert retrieval. In the target-present JOL condition, a category name was provided; thus, it was unlikely to elicit covert retrieval while making JOLs. If covert retrieval plays an important role in JOL reactivity, it is anticipated that the target-absent JOL group will outperform the restudy control group, whereas the target-present JOL group will not. This prediction aligns with the covert retrieval hypothesis. However, if covert retrieval does not suffice to enhance inductive learning, neither the target-absent nor target-present JOL groups will surpass the control group. In Experiment 2, we directly tested the effect of making JOLs framed to afford overt versus covert retrieval attempts by comparing the retrieval practice + JOL condition to the target-absent JOL condition. If learners do not spontaneously engage in active retrieval when they are not explicitly asked to do, it is anticipated that the retrieval practice + JOL group will outperform the other groups, as predicted by truncated search hypothesis. Experiment 3 further investigated whether the act of making JOLs added any additional learning benefit to mere overt retrieval practice (without JOL) by comparing the retrieval practice + JOL condition to the retrieval practice condition. If making JOLs facilitates learning in ways distinct from retrieval, it is expected that the retrieval practice + JOL group will outperform the retrieval practice group. However, if the direct benefit of making JOLs is not significant (if any) and/or its benefit is redundant to the benefit of retrieval practice, the performance

difference between the two groups will not be pronounced. Nonetheless, we predict that both the retrieval practice + JOL and retrieval practice groups will outperform the restudy control group, given the robust effect of retrieval practice demonstrated in inductive learning (e.g., Jacoby et al., 2010; Kang et al., 2023).

Experiment 1

Experiment 1 investigated whether making JOLs of prior learning could facilitate the learning of old and new materials in inductive learning. We also varied the form of the JOL (target-absent JOL or target-present JOL) to examine the role of covert retrieval. If the target-absent JOL elicited covert retrieval, this JOL group would outperform the target-present JOL and restudy groups.

Method

Participants

To determine the sample size, power analysis was performed using G*Power (Faul et al., 2007). Previous studies on JOL reactivity (Carvalho et al., 2023; Kubik et al., 2022; Putnam & Roediger, 2013) reported medium to very large effect sizes ranged from 0.45 to 1.06 (Cohen's d). To detect a large-sized effect ($\eta_p^2 = 0.14$) among the three between-subjects conditions, at least a total sample size of 81 was required at a power of 0.9. Thus, for each experiment, we recruited 30 or more participants per condition, excluding those who did not meet the predefined exclusion criteria. In Experiment 1, 95 undergraduates from a large university in South Korea were recruited and tested online in return for monetary compensation. We excluded two participants who were inactive for more than 10 min on the self-paced page, leaving 93 participants (60 women; 33 men; mean age = 22.37 years) for the final analyses. Each participant was randomly assigned to one of the three conditions: target-absent JOL (31 participants), target-present JOL (32 participants), and interim-restudy (30 participants). The Institutional Review Board (IRB) of the university where the research was conducted approved all experiments of the current study.

Design

A 3 (interim activity: target-absent JOL vs. target-present JOL vs. interim-restudy) $\times$ 2 (section: A vs. B) mixed design was employed. The type of interim activity was manipulated as a between-subjects factor and the section was a within-subject factor. According to the randomly assigned conditions, participants were instructed to make either target-absent JOLs, target-present JOLs, or did an interim restudy of the materials in Section A.

Materials

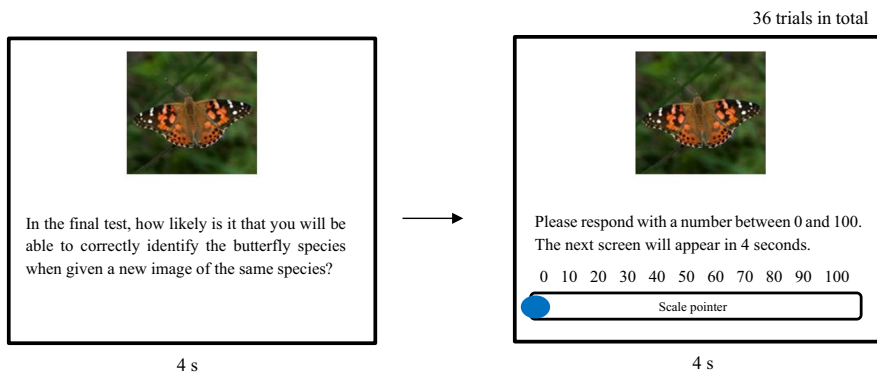
To investigate the effect of making JOLs on inductive learning, we selected butterfly species. Butterfly images were initially created by Birnbaum et al. (2013), and some of the images were replaced with those found in the Korean Encyclopedia of Butterflies to adjust for the difficulty of the learning materials. There were 12 butterfly species with each species consisting of six photos (72 images in total): *Apatura iris*, *Dichorragia nesimachus*, *Libythea lepita*, *Lobocla bifasciata*, *Lycaenidae*, *Minois dryas*, *Neptis philyra*, *Nymphalis xanthomelas*, *Parnassius stubbendorffii*, *Polygonia c-aureum*, *Ypthima baldus*, and *Vanessa cardui*. All the materials, including instructions and butterfly species names, were prepared and presented in Korean. The butterfly species were referred to by their common names in Korean, rather than their technical binomial names.

Procedure

Participants were tested online using Qualtrics (<https://www.qualtrics.com/>). Participants initially received general instructions regarding the study purpose and procedure. They were instructed to study 12 different butterfly species across two sections (Sections A and B) with six species per section. To control for general test expectancy, all participants were told that they would take an examination requiring them to classify new butterfly images from the studied butterfly species.

Participants studied six of the 12 butterfly species in Section A and the other six species in Section B. To avoid the specific-item effect, the butterfly-section pairs were counterbalanced across the two sections. In Section A, 36 butterfly images (six species $\times$ six images) were individually presented. For each trial, an image and the corresponding species name were concurrently shown in the middle of a computer screen for 4 s. After studying the Section A materials, participants were asked to perform one of three interim activities with the same set of butterfly images presented in Section A. In both the target-absent and target-present JOL conditions, participants were asked to make a JOL. Figure 2 shows illustrations of the JOL procedure employed in Experiment 1. The JOL was conducted across two separate pages, with each page displayed for 4 s, totaling 8 s. Participants were not allowed to provide a JOL rating during the first 4 s to ensure adequate processing of the presented image. On the first page, an image of butterfly was presented with a JOL question. Specifically, participants were asked to predict the likelihood of correctly identifying the butterfly species when given a new image of the same species in a final test. On the second page, participants were asked to give a rating by moving the scale pointer from 0 (not at all) to 100 (definitely). The only difference between the target-absent and target-present conditions was the absence or presence of butterfly species names. In the target-absent JOL condition (Fig. 2a), only the cue (i.e., butterfly image) was presented whereas in the target-present JOL condition (Fig. 2b), both the cue (i.e., butterfly image) and target (i.e., butterfly species name) were

(a) target-absent JOL condition



(b) target-present JOL condition

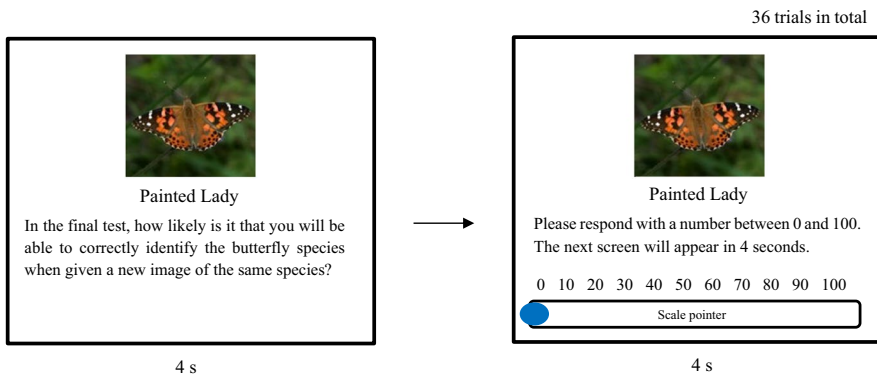


Fig. 2 Illustrations of JOL procedure used in Experiment 1. The butterfly image was obtained from Birnbaum et al. (2013) and then adapted for the current study. All instructions and butterfly species were provided in Korean

simultaneously presented while making JOLs. In the interim-restudy condition, participants were asked to restudy the identical image sets they studied in Section A; thus, they did not make any JOLs. Each image was presented for 8 s. After completing the interim session in Section A, all participants studied the six new butterfly species in Section B under the same conditions. As in Section A, there were 36 butterfly images (six species $\times$ six images), and each image was presented individually for 4 s. All participants then took a final transfer test which included four new butterfly images from each of the 12 butterfly species studied, comprising 48 test items. In each test trial, a butterfly image was presented, and the participants had to choose the corresponding butterfly species from the 12 different names (Sections A and B) presented in alphabetical order. The final test was self-paced and there was no feedback. The experiment lasted approximately 30 min and participants were debriefed at the end.

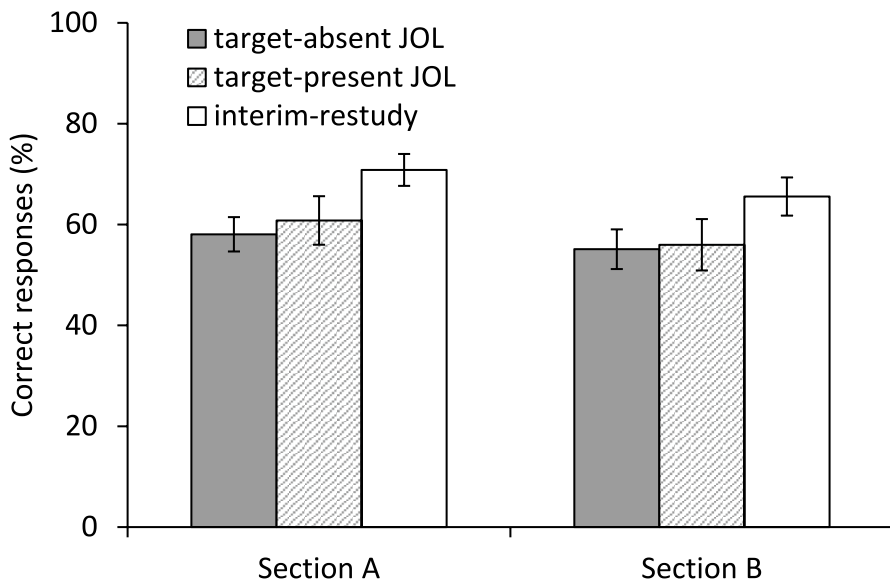


Fig. 3 Mean percentages of correct responses in the final transfer tests of Sections A and B for the target-absent JOL, target-present JOL, and interim-restudy conditions of Experiment 1. Error bars represent one standard error of the mean

Results

Interim activity performance

Participants in both JOL conditions provided JOL ratings for the six butterfly species studied in Section A. The mean JOL rating was 70.12 ($SD = 22.67$) in the target-absent JOL condition and 65.86 ($SD = 20.70$) in the target-present JOL condition. There was no significant difference in ratings between the two JOL groups, $t(61) = 0.78$, $p = .438$.

Final test performance

For analyses of final test performance, we reported statistics for both traditional null hypothesis significance testing and Bayes Factors (BF). We reported BF_{10} when p-values were significant and BF_{01} when p-values were not significant so that a larger value of BF indicated stronger support for the reported effect. According to Jeffreys' scheme (Jeffreys, 1961), the Bayes factor is interpreted as follows: 1 to 3 provides anecdotal evidence, 3 to 10 indicates moderate evidence, 10 to 30 suggests strong evidence, 30 to 100 provides very strong evidence, and 100 or more provides decisive evidence. These analyses were performed using JASP (JASP Team, 2024).

Figure 3 shows the mean percentages of correct responses in the final tests in Sections A and B. A 3×2 mixed analysis of variance (ANOVA) was performed on the number of correct answers in the final test. Interim activity (target-absent JOL vs. target-present JOL vs. interim-restudy) was a between-subjects factor, and section (A vs. B) was a within-subject factor.

There was neither main effect of the interim activity, $F(2, 90) = 2.88, p = .061, BF_{01} = 0.85$, nor was there an effect of the section, $F(1, 90) = 3.77, p = .055, BF_{01} = 1.10$. Also, no interaction occurred between the section and interim activity, $F(2, 90) = 0.10, p = .905, BF_{01} = 9.04$.

Discussion

Experiment 1 examined the effect of making JOLs on inductive learning and obtained the same patterns of results across the two sections. Regarding transfer performance in Sections A and B, all three conditions showed similar levels of learning performance, indicating no JOL reactivity in inductive learning. Neither target-absent nor target-present JOL was more effective than the interim-restudy in improving the learning of previously and newly learned butterfly species. We also failed to observe a performance difference between the target-absent and target-present JOL conditions. These results are inconsistent with those of the previous studies that demonstrated a positive forward effect (Kubik et al., 2022) and backward effect of JOLs (Dougherty et al., 2005; Soderstrom et al., 2015; Witherby & Tauber, 2017) in verbal learning. One possible explanation for the null effect of JOL is that inductive learning is more difficult and complex than verbal learning. Most previous studies reporting positive JOL reactivity used retention tests with simple learning materials such as word pairs (Halamish & Undorf, 2022; Tauber & Witherby, 2019) and word lists (Li et al., 2021; Li et al., 2023; Zhao et al., 2022). In contrast, the current experiment used category learning materials for a transfer test in which participants had to apply what they studied to new exemplars. Such transfer tasks require not only memorization of category names but also generalization of learning. Thus, even if covert retrieval was elicited while making JOLs in the target-absent condition, it may have been insufficient to achieve a level of learning that went beyond memorization. Consistent with this explanation, Ariel et al. (2021) did not find benefits of target-absent JOLs compared with restudy when using a text comprehension task, which requires a relatively higher level of learning than simple verbal learning.

Moreover, the null reactivity from both the current experiment and Ariel et al.'s research (2021) suggests that students do not spontaneously engage in active retrieval attempts, and the process of covert retrieval is probably different from overt retrieval practice. While making JOLs, even when the target (i.e., butterfly species name) was not presented, participants may have produced a number rating without retrieving the target item. Because participants were not explicitly instructed to retrieve the butterfly species name, they did not have to recall it; thus, they did not benefit from the retrieval practice. In addition, no retrieval attempts likely failed to evoke any retrieval failure, a phenomenon known to enhance learning (Cho et al., 2017). Indeed, learners who make JOLs are not necessarily fully engaged in the

retrieval process (Tekin & Roediger, 2021). This was because learners were more likely to make judgments based on familiarity cues and tended to terminate the retrieval process too early, which is consistent with the truncated search hypothesis (Tauber et al., 2015). Dobson et al. (2019) also found that the JOL group spent less time on the retrieval process than the retrieval practice group. These results suggest that overt retrieval may be necessary for the act of making JOLs to produce learning gains.

In addition, because the procedure for making JOLs was conducted at the experimenter's pace in Experiment 1, the participants might not have had enough time for a retrieval attempt. Under the constraint of a time limit, participants may have been preoccupied with selecting numbers on a scale instead of trying to retrieve the learned materials. Learners are expected to covertly retrieve items before producing a JOL (Jönsson et al., 2012); however, the covert retrieval process may require sufficient time. Making self-paced JOLs may allow learners to fully reflect on their learning status and induce a more active retrieval process. To address this possibility, Experiment 2 asked participants to make a target-absent JOL at their own pace, and directly compared its effect to making JOLs with overt retrieval.

Experiment 2

Experiment 2 investigated whether making JOLs with overt retrieval would enhance the learning of previously and newly learned materials in inductive learning. Three modifications were made from Experiment 1. First, to allow sufficient time for retrieval, participants were instructed to make JOLs in a self-paced manner. Second, we removed the target-present JOL condition because it did not exhibit any learning benefits in Experiment 1. Third, we added a new condition, retrieval practice + JOL, and compared its effect with the target-absent JOL condition, in addition to restudy control, to determine the effect of overt versus covert retrieval. If overt retrieval opportunities play an important role in JOL reactivity, the retrieval practice + JOL group would outperform the target-present JOL and restudy groups.

Method

Participants

In total, 108 undergraduates from a large university in South Korea were recruited and tested online in return for monetary compensation. We excluded two participants who were inactive for more than 10 min on the self-paced page, one whose interim test score was below the three SDs of the group mean, and two who participated twice in duplicate, leaving 103 participants (63 women; 40 men; mean age = 22.69 years) for the final analyses. Each participant was randomly assigned to one of the three conditions: retrieval practice + JOL (35 participants), target-absent JOL (34 participants), or interim-restudy (34 participants).

Design, Materials, and Procedure

The biggest change in Experiment 2 was the removal of the target-present JOL condition and the addition of a new condition: retrieval practice + JOL. In the retrieval practice + JOL condition, the retrieval practice and JOL procedures were administered across two separate pages. On the first page, participants were presented with a butterfly image and prompted to type in the corresponding butterfly species name. There was no time limit or feedback for any trial. On the second page, the butterfly image disappeared from the screen and participants were presented with a JOL question. They were asked to indicate how likely it was they would be able to correctly identify the butterfly species they had just viewed when given a new image of the same species in a final test by entering a number between 0 and 100. The procedure for the target-absent JOL condition was identical to that of the retrieval practice + JOL condition, with one exception: participants were not prompted to enter the butterfly species name and instead simply made a JOL. Unlike Experiment 1, the JOL procedure in Experiment 2 was conducted at the participants' own pace, allowing them to take as much time as necessary to assess their learning status. The restudy control condition and all other procedures were the same as those used in Experiment 1.

Results

Interim Activity Performance

Participants in both JOL conditions provided JOL ratings for the six butterfly species studied in Section A. The mean JOL rating was 66.48 ($SD = 19.85$) in the retrieval practice + JOL condition and 74.82 ($SD = 16.80$) in the target-absent JOL condition. There was no reliable rating difference between the two interim JOL groups, $t(67) = -1.88, p = .064$.

As participants made JOLs in a self-paced manner, we also examined whether the time spent to make JOLs differed between the two JOL conditions. An independent t -test revealed a significant difference, $t(67) = 10.76, p < .001$, such that the retrieval practice + JOL group ($M = 8.71, SD = 3.29$) spent a significantly longer time than the target-absent JOL group ($M = 2.50, SD = 0.71$). The mean percentage of correct responses in the retrieval practice + JOL condition was 68.57 ($SD = 18.11$).

Final Test Performance

Figure 4 shows the mean percentage of correct responses in the final tests of Sections A and B in the interim JOLs and interim-restudy conditions. A 3×2 mixed ANOVA was performed on the number of correct answers. Interim activity (retrieval practice + JOL vs. target-absent JOL vs. interim-restudy) was a between-subjects factor, and section (A vs. B) was a within-subject factor.

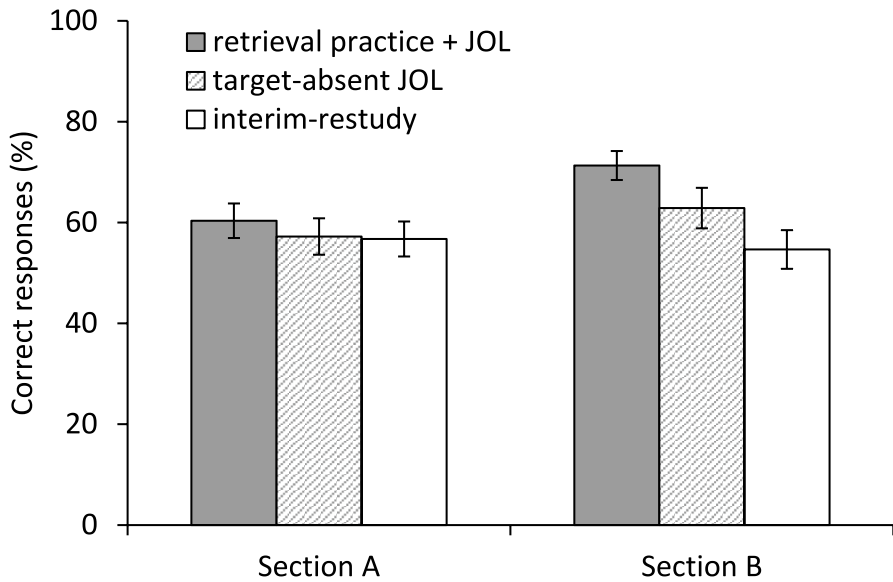


Fig. 4 Mean percentages of correct responses in the final transfer tests of Sections A and B in the retrieval practice + JOL, target-absent JOL, and interim-restudy conditions of Experiment 2. Error bars represent one standard error of the mean

There was no main effect of interim activity, $F(2, 100) = 2.80$, $p = .066$, $BF_{01} = 1.06$. There was a significant main effect of the section, $F(1, 100) = 5.32$, $p = .023$, $\eta_p^2 = .05$, $BF_{10} = 1.76$, indicating that the transfer performance was significantly higher in Section B ($M = 63.03$, $SD = 21.97$) than in Section A ($M = 58.13$, $SD = 20.38$).

Notably, there was a significant interaction between the interim activity and section, $F(2, 100) = 3.26$, $p = .042$, $\eta_p^2 = .06$, $BF_{10} = 2.02$, suggesting that the effect of the interim activity varied depending on the section. Different patterns of results were observed in Sections A and B. For all pairwise comparisons, we conducted a Bonferroni adjustment to maintain a family-wise Type I error rate of .05. The follow-up one-way ANOVAs revealed no group differences in Section A, $F(2, 100) = 0.32$, $p = .729$, $BF_{01} = 8.64$, but a significant group difference in Section B, $F(2, 100) = 5.38$, $p = .006$, $\eta_p^2 = .10$, $BF_{10} = 6.55$. Specifically, only the retrieval practice + JOL group outperformed the interim-restudy group, $MD = 16.65$, 95% CI [4.58, 28.73], $p = .004$, $BF_{10} = 34.71$. The target-absent JOL group did not differ significantly from the retrieval practice + JOL group, $MD = 8.44$, 95% CI [-3.64, 20.52], $p = .298$, $BF_{01} = 1.16$, or the interim-restudy group, $MD = 8.21$, 95% CI [-3.95, 20.37], $p = .334$, $BF_{01} = 1.59$.

Discussion

Experiment 2 examined the effects of making JOLs designed to evoke covert retrieval and making JOLs that were accompanied by overt retrieval practices compared to a restudy control. The results revealed that the effects of making JOLs differed depending on the section. For Section A, there was no reliable performance difference among the three conditions, replicating the findings of Experiment 1, in that there was no backward effect of JOLs. Further, Experiment 2 found that even when JOL was accompanied by overt retrieval practice, it did not enhance learning of the judged material¹. This null effect is different from the results of a previous study demonstrating that making JOLs with retrieval practices on previously learned texts enhanced the learning of the studied text (Ariel et al., 2021).

In Section B, making JOLs enhanced learning, but only when learners were asked to retrieve information prior to making JOLs. That is, retrieval practice + JOL showed the forward effect of JOL, whereas target-absent JOL did not. Interestingly, all participants studied in Section B in the same manner. They simply engaged in different interim activities with regard to Section A before moving on to Section B but showed different levels of new learning. One possible explanation is that because the retrieval practice + JOL group had to overtly retrieve the category name, they perhaps experienced more retrieval failure and, in turn, exerted more effort into their learning in the subsequent section (Cho et al., 2017; Potts & Shanks, 2014; Yang et al., 2017). By contrast, the target-absent JOL group did not have to recall the category name; thus, they were less likely to engage in retrieval attempts and experience retrieval failure. Indeed, the target-absent JOL group spent significantly less time than the retrieval practice + JOL group, lending support to the idea that the former group exerted less effort on retrieval than the latter. These results suggest that, when making JOLs, performing overt retrieval practice is important to facilitate the learning of new materials.

If making JOLs enhances learning only when combined with overt retrieval, the following question is whether making JOLs provides additional learning benefits to retrieval practice. Retrieval practice may be good enough by itself without the additional act of making JOLs, especially given that making JOLs requires learners to spend additional time. Previous studies have shown that retrieval fosters learning (Carpenter, 2012; Roediger & Karpicke, 2006), and such retrieval benefits have also been found in inductive learning as well (e.g., Jacoby et al., 2010; Lee & Ahn, 2018;

¹ To investigate whether the backward effects varied based on retrieval success during the interim session, we analyzed the performance data by butterfly species. While some species (such as *Parnassius stubbendorffii* and *Ypthima baldus* in Experiment 2 and *Minois dryas*, *Parnassius stubbendorffii*, and *Ypthima baldus* in Experiment 3) exhibited significantly higher retrieval success than others ($ps < .01$), none of the butterfly species in Experiments 2 and 3 demonstrated backward effects of retrieval practice. Even when retrieval success was high for a given species, neither the retrieval + JOL group (Experiments 2 and 3) nor the retrieval practice group (Experiment 3) outperformed the restudy control group on the final transfer test of Section A. This was because the restudy group performed as well as other groups for categories with high retrieval success during the interim session, suggesting varying difficulty levels among different butterfly species. Despite such varying levels of difficulty among butterfly species, the lack of the backward effect remained consistent across different categories.

Yang & Shanks, 2018). Although the retrieval practice + JOL group showed better transfer performance in Section B than did the interim-restudy group, it remained unclear whether the benefit was entirely due to retrieval. Thus, Experiment 3 examined the comparative effectiveness of retrieval practices alone versus retrieval practices combined with JOL.

Experiment 3

Experiment 2 revealed the act of making JOLs potentiated new learning only when overt retrieval practice was conducted before providing JOLs. Experiment 3 investigated whether the act of making JOLs additionally benefits a mere overt retrieval practice by comparing the retrieval practice + JOL and retrieval practice conditions. Two modifications were made from Experiment 2. First, we removed the target-absent JOL condition and added a retrieval practice condition in which participants were instructed to make only overt retrievals without making JOLs. Second, to increase the external validity, we broadened the participant pool by recruiting undergraduate students from multiple universities in South Korea.

Method

Participants

A total of 113 undergraduates from three private universities located in Seoul, South Korea, were recruited and tested online in return for monetary compensation. We excluded two participants who were inactive for more than 10 min on the self-paced page, one whose interim test scores were below the three SDs of the group mean, two who reported taking notes, and three who participated twice in duplicate, leaving 105 participants (77 women; 28 men; mean age = 22.05 years) for the final analyses. Each participant was randomly assigned to one of the three conditions: retrieval practice + JOL (35 participants), retrieval practice (35 participants), or interim-restudy (35 participants).

Design, Materials, and Procedure

The overall procedure used in Experiment 3 was identical to that used in Experiment 2. Participants performed either overt retrieval practice prior to making JOLs (retrieval practice + JOL condition), overt retrieval practice (retrieval practice condition), or restudied (interim-restudy condition) on Section A materials before continuing with Section B. In the retrieval practice condition, participants did not provide any judgments and only performed the retrieval practice. They were presented with a butterfly image and instructed to enter a category name during the interim session. There was no feedback or time limit for any trial. The other two conditions were identical to those used in Experiment 2.

Results

Interim Activity Performance

Only the retrieval practice + JOL group provided judgments in Section A. The mean rating was 73.50 ($SD = 16.59$).

We also examined the retrieval practice performance of the two conditions: retrieval practice + JOL and retrieval practice. An independent t -test did not reveal a significant mean difference, $t(68) = -0.29$, $p = .769$, between the retrieval practice + JOL ($M = 73.33$, $SD = 17.53$) and retrieval practice ($M = 74.68$, $SD = 20.62$) conditions. However, an independent t -test revealed a significant difference in the duration of the interim session, $t(68) = 3.79$, $p < .001$, in that the retrieval practice + JOL group ($M = 9.28$, $SD = 3.85$) spent more time than the retrieval practice group ($M = 5.85$, $SD = 3.75$).

Final Test Performance

Figure 5 shows the mean percentage of correct responses in the final test of Sections A and B in the retrieval practice + JOL, retrieval practice, and interim-restudy conditions. A 3×2 mixed ANOVA was performed on the number of correct answers. Interim activity (retrieval practice + JOL vs. retrieval practice vs. interim-restudy) was a between-subjects factor, and section (A vs. B) was a within-subject factor.

There was a significant main effect of the interim activity, $F(2, 102) = 4.00$, $p = .021$, $\eta_p^2 = .073$, $BF_{10} = 2.49$. The retrieval practice + JOL group outperformed the interim-restudy group, $MD = 11.49$, 95% CI [1.21, 21.76], $p = .023$, $BF_{10} = 14.55$. No other intergroup differences were observed. Also, there was no main effect of the section, $F(1, 102) = 0.80$, $p = .375$, $BF_{01} = 4.59$.

Interestingly, there was a significant interaction between the interim activity and section, $F(2, 102) = 4.77$, $p = .011$, $\eta_p^2 = .09$, $BF_{10} = 2.11$, suggesting that the effect of the interim activity varied depending on the section. Different patterns of results were observed in the two sections. The follow-up one-way ANOVAs revealed no group differences in Section A, $F(2, 102) = 0.66$, $p = .520$, $BF_{01} = 6.66$, whereas for Section B, there were significant group differences, $F(2, 102) = 7.39$, $p = .001$, $\eta_p^2 = .13$, $BF_{10} = 31.03$. Specifically, the retrieval practice + JOL group outperformed the interim-restudy group, $MD = 17.62$, 95% CI [5.79, 29.45], $p = .002$, $BF_{10} = 16.38$. The retrieval practice group also outperformed the interim-restudy group, $MD = 15.24$, 95% CI [3.41, 27.07], $p = .008$, $BF_{10} = 11.07$. However, the retrieval practice group did not significantly differ from the retrieval practice + JOL group, $MD = 2.38$, 95% CI [-9.45, 14.21], $p = 1.00$, $BF_{01} = 3.57$.

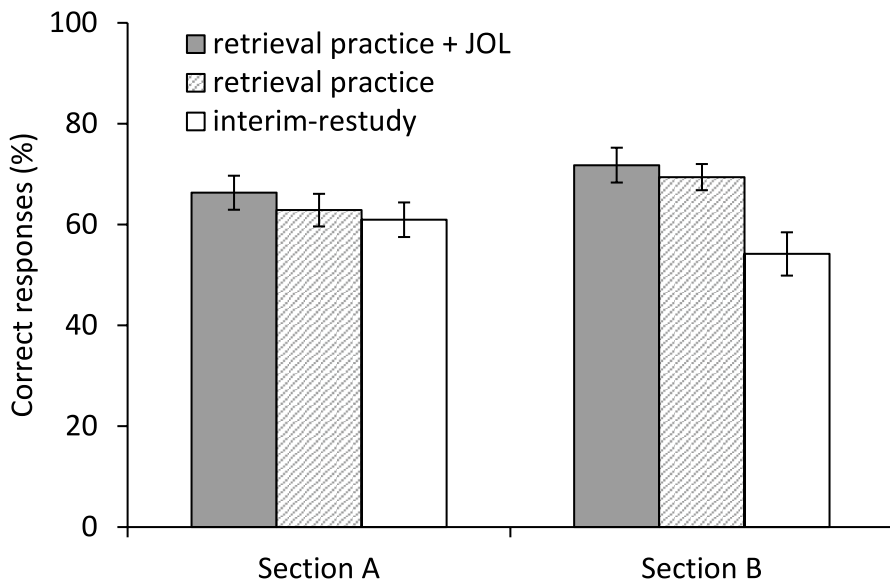


Fig. 5 Mean percentages of correct responses in the final transfer tests of Sections A and B in the retrieval practice + JOL, retrieval practice, and interim-restudy conditions of Experiment 3. Error bars represent one standard error of the mean

Discussion

Experiment 3 investigated the effect of combining JOLs with retrieval practice on inductive learning by comparing these effects to those of retrieval practice alone and interim-restudy conditions. For Section A, there was no reliable performance difference among the three groups, replicating the findings of Experiment 2 in that there was no backward effect of JOLs compared with the interim-restudy condition. In Section B, both the retrieval practice + JOL and retrieval practice groups outperformed the interim-restudy group, indicating a forward effect. The results of the retrieval practice + JOL group replicated those of Experiment 2. Performing retrieval practice prior to making JOLs enhanced new learning compared to the interim-study.

However, both the retrieval practice + JOL and retrieval practice groups exhibited a comparable level of transfer performance in Section B. Making JOLs accompanied by retrieval practice was as effective as retrieval practice alone in enhancing the learning of new materials, but was not more effective than retrieval practice alone. That is, making JOLs in addition to retrieval practices did not provide any additional learning benefits. These results suggest that the forward effect of retrieval practice + JOL observed in Experiment 2 was probably due to overt retrieval alone rather than the combined effect of JOL and the retrieval practice. The observed forward effect of retrieval practice is consistent with the findings of prior research demonstrating the forward effect of testing (for reviews, see Chan et al., 2018b; Pastötter & Bäuml, 2014; Yang et al., 2018). Several studies have specifically shown that testing

on previously learned materials can potentiate new learning in the field of inductive learning by using natural rock categories (Kang et al., 2023) and painting styles (Lee & Ahn, 2018; Yang & Shanks, 2018).

Notably, the retrieval practice + JOL group spent significantly more time during the interim session than the retrieval practice group, suggesting that retrieval practice is a more efficient learning method. Our findings indicate that retrieval may play a crucial role in improving inductive learning.

General Discussion

The present study examined whether making JOLs enhances inductive learning and how their effects change depending on the extent to which JOLs afforded retrieval attempts. Across the three experiments, the form of JOL was manipulated, and all JOL conditions were compared with the restudy control condition to examine the forward and backward effects of JOLs. Specifically, Experiment 1 tested the effects of making target-absent and target-present JOLs. Experiment 2 examined the effects of JOLs accompanied by overt retrieval practice and target-absent JOLs. Finally, Experiment 3 examined the effects of JOLs accompanied by overt retrieval and mere retrieval practice. These manipulations allowed us to investigate the roles of both covert and overt retrieval attempts in JOL reactivity during inductive learning.

Our results showed that standard forms of JOLs, either target-absent or target-present JOLs, did not enhance inductive learning. Participants who made target-absent JOLs (Experiments 1 and 2) or target-present JOLs (Experiment 1) did not show better transfer performance in the final test of both Sections A and B compared to the restudy control group, indicating no backward and forward effects of JOLs. However, when overt retrieval practice was prompted prior to making JOLs (Experiments 2 and 3), the subsequent learning of new material was enhanced. Participants who made overt retrieval + JOL exhibited better transfer performance in the final test of Section B than the restudy control group, indicating a forward effect of JOLs. Such a forward effect was also observed when participants performed retrieval practice only without making JOLs (Experiment 3), showing a forward effect of testing.

The present findings suggest that the mere act of making JOLs may not be sufficient to enhance inductive learning. Both target-absent and target-present JOLs failed to show backward or forward effects when they were not accompanied by overt retrieval practice. Our results are inconsistent with those of many prior studies that have demonstrated positive JOL reactivity with simple verbal learning materials (e.g., Kubik et al., 2022; Soderstrom et al., 2015). There are three possible explanations for the lack of positive JOL reactivity in the present study. First, JOL reactivity may depend on the type of learning material involved. Most previous studies on positive JOL reactivity have been conducted using verbal materials, such as word pairs and word lists (e.g., Janes et al., 2018; Soderstrom et al., 2015; see also Double et al., 2018 for a review). Making JOLs may be sufficient to enhance retention of simple materials. However, the current study asked participants to learn about various butterfly species; thus, they had to abstract and remember the species characteristics from the studied exemplars. Making JOLs may not be sufficient to enhance

learning, which requires both the abstraction and retention of complex learning materials such as natural categories. Exposure to a cue (in the target-absent JOL condition) or cue-target (in the target-present JOL condition) may not necessarily help inductive learning. Consistent with this account, Ariel et al. (2021) did not find positive effects of eliciting standard forms of JOLs on text comprehension, suggesting that the direct effect of JOLs varies depending on the complexity of the learning materials and the type of learning goal. In the current study, higher-order goals such as categorization may have presented a higher cognitive burden for learners. Cognitive load theory (CLT) suggests that learners' cognitive resources are limited, and high working memory demand can hinder learning (Sweller, 1994; Sweller et al., 2019). As natural category learning deals with ambiguous and unclear characteristics between categories (Miyatsu et al., 2019), both studying categories and making JOLs might have made learners consume cognitive resources at a significant level. Hence, simply being exposed to a cue (in the target-absent JOL condition) or a cue-target (in the target-present JOL condition) while making JOLs did not necessarily enhance learning.

Second, participants may not have fully engaged in the covert retrieval process while making JOLs. In this study, participants who made target-present JOLs did not have to retrieve any information because they were provided with both cues (butterfly image) and targets (butterfly species name). Thus, it is not surprising that the target-present JOL group did not outperform the restudy control group. This finding is consistent with previous studies that reported a lack of a direct effect of target-present JOLs. For example, Schäfer and Undorf (2024) found that participants who made target-present JOLs did not show improved performance in terms of general knowledge. Kubik et al. (2022) also demonstrated that making JOLs with complete information did not enhance the subsequent learning of word lists. However, in our study, even target-absent JOLs did not show any learning benefit compared to the restudy control. Although the participants in the target-absent JOL condition were expected to covertly retrieve the items before judging (Jönsson et al., 2012; Spellman & Bjork, 1992), they did not have to do so because they were simply required to generate a number to judge their learning status. Therefore, if participants did not spontaneously attempt to retrieve the butterfly species name, they would not benefit from making target-absent JOLs. Previous studies have reported that learners who make JOLs are not necessarily fully engaged in the retrieval process (Tekin & Roediger, 2021), suggesting that they sometimes complete the retrieval process prematurely (Son & Metcalfe, 2000; Tauber et al., 2015). Dobson et al. (2019) also reported that participants who made JOLs spent less time on the retrieval process than those who had overt retrieval practice.

Finally, it is possible that the backward effect of JOLs did not occur because the final test was a transfer task. The current study used a transfer task, instead of a retention task, to examine the backward and forward effects of making JOLs on inductive learning. This is because the goal of inductive learning is not just to remember specific cases, but also to generalize learning to other instances (Holland et al., 1989). According to Soderstrom et al. (2015), one of the conditions under which JOL reactivity occurs is that the same cues used in the JOL tasks must be available in the final criterion test. Therefore, for JOL reactivity to occur, the

photograph of a butterfly presented in the JOL task should also be presented in the final test. However, the final test in the current study asked participants to classify new exemplars that were not seen in any of the previous learning sessions. Future studies should examine how JOL reactivity varies according to cue availability and test type in inductive learning.

Although eliciting JOLs in typical forms did not facilitate inductive learning, we repeatedly found across Experiments 2 and 3 that JOLs accompanied by retrieval practice enhanced inductive learning. When participants were prompted to make overt retrievals before making JOLs, they outperformed the restudy group on the final transfer test of Section B, showing a forward effect of JOLs. However, retrieval practice + JOL was not more effective than, but was as effective as retrieval practice alone, suggesting that retrieval practice played an important role in facilitating subsequent inductive learning. We conjecture that performing overt retrieval before making JOLs may have worked similarly to interim testing. Numerous studies have demonstrated that taking an interim test on previously learned materials can enhance subsequent learning (for a review, see Yang et al., 2018), a phenomenon often referred to as the forward effect of testing (Pastötter & Bäuml, 2014), interim test effect (Wissman et al., 2011), or test-potentiated new learning (Chan et al., 2018a). The benefits of interim testing have also been shown in inductive learning, encompassing materials such as painting styles (Lee & Ahn, 2018; Yang & Shanks, 2018) and rock categories (Kang et al., 2023). For example, Kang et al. (2023) had undergraduate students learn various rock categories and demonstrated that interim testing on previously studied rock categories facilitated the learning of both previously studied and newly studied categories compared to restudy. In the present study, although participants in the retrieval practice + JOL condition were not given a formal test, the prompt to enter the name of the butterfly species in every trial before making a JOL was more or less the same as taking an interim test. This retrieval process could have enhanced new learning by affecting study strategies for the encoding and absorption of subsequent materials (Chan et al., 2018a; Cho et al., 2017; Finley & Benjamin, 2012; Pyc & Rawson, 2012). When participants attempted retrieval prior to making JOLs, they could evaluate their current learning strategies and adjust them to be more effective in subsequent learning. In addition, they may have experienced retrieval failure of some butterfly species while making JOLs and exerted more effort in their subsequent learning of new species. Retrieval practice is known to increase learners' motivation (Cho et al., 2017) and decrease mind-wandering during subsequent learning (Szpunar et al., 2013).

However, it should be noted that retrieval practice + JOL was not more effective than retrieval practice alone. Experiment 3 compared these two conditions to restudy control to investigate whether the benefit of retrieval practice + JOL was due to the combined effect of JOL and retrieval practice, or due to retrieval practice itself. The results revealed no additional benefits from making JOLs. In fact, making JOLs accompanied by retrieval practice was less efficient than mere retrieval practice in that the former required a longer time than the latter. These results are consistent with several previous studies demonstrating testing benefits in natural category learning (e.g., Jacoby et al., 2010; Kang et al., 2023) and add further evidence that retrieval practice is an effective learning tool (Roediger & Butler, 2011).

According to our findings, having students perform retrieval practice may be sufficient to facilitate inductive learning. Even though all participants in Experiment 3 studied the same materials for the same amount of time in Section B, participants who had retrieval practice for Section A materials showed better learning for Section B materials.

Nonetheless, our findings do not suggest that making JOLs does not do any good for learning. It is possible that there was a hidden efficacy not revealed in the current study. For example, making JOLs may help learners optimize their studies in self-regulated learning. The current study provided participants with fixed learning sessions for both Sections A and B to control for exposure to study materials across the conditions (we implemented a self-paced interim JOL session in Experiments 2 and 3, but each study section was always experimenter-paced). If students were allowed to regulate their studies in Section B, those who made interim JOLs for earlier learning might have been able to allocate study time more effectively in their subsequent studies. Indeed, prior research has shown that people use metacognitive judgments to regulate their subsequent study behaviors, such as study time allocation (Son & Metcalfe, 2000) and the selection of restudy materials (Morehead et al., 2016). Future research should investigate how JOL reactivity changes depending on the extent to which students can regulate their studies.

Furthermore, JOL reactivity can change depending on the item-relatedness of the cue and target (i.e., how strongly a cue is related to a target). One explanation for positive JOL reactivity is the cue-strengthening hypothesis (Soderstrom et al., 2015). According to this hypothesis, while making JOLs for related word pairs, the semantic relationship between the two words serves as a strong cue that reinforces memory traces. In contrast, unrelated word pairs have little meaningful relationships and do not provide a strong retrieval cue. Furthermore, the relatedness processing of unrelated cue-targets may induce dual-task costs that impede learning (Mitchum et al., 2016). For this reason, no reactivity or negative reactivity has been shown for unrelated pairs (Double et al., 2018). Likewise, in the present study, item relatedness of the cue and target (butterfly exemplar – species) might have affected the results. In our study, however, the cue-target can be seen always “related” because a butterfly exemplar always belongs to a specific species, thus making it hard to analyze the moderating effect of item relatedness as conducted in previous studies. Alternatively, we could use the similarity between the study and transfer items instead, but unfortunately, the items used in our study do not have a wide range of similarities. Because all butterfly species were unfamiliar to participants (note that unlike word pairs, participants had to induce category characteristics for themselves and make a link between those characteristics and category label), unusual/atypical exemplars were not included in experiments. Future studies need to manipulate typicality of exemplars (if an exemplar is very different from other exemplars, it can be considered to have weak item relatedness) or item similarity between study and transfer items to see how JOL reactivity varies depending on various aspect of learning materials.

We did not find backward effects of making JOLs or performing retrieval practices in any of the reported experiments. The retrieval practice + JOL and retrieval practice groups did not outperform the restudy group in the final transfer test of

Section A. This result may appear unexpected given the large volume of research showing the typical backward testing effect. Furthermore, this is inconsistent with a previous study that demonstrated the benefits of making JOLs accompanied by retrieval in text comprehension (Ariel et al., 2021). More recently, however, Zhao et al. (2023) reported a null effect of making JOLs accompanied by overt retrieval in text learning. There could be several explanations as to why retrieval did not benefit inductive learning in the present study. First, the learners did not have restudy opportunities or receive corrective feedback after retrieval practice. Performing retrieval practice without feedback might be a disadvantage compared to the interim-restudy condition, in which participants were always provided with both butterfly images and species names. Retrieval practice effectiveness increases when learners can correct errors through restudy (Amlund et al., 1986) and when explicit feedback is provided (Kang et al., 2007; Pashler et al., 2005). Second, the final test of the current study was an immediate test that challenged the appearance of backward testing effect. Previous studies found that testing effects were often greater when tests were delayed (e.g., Karpicke & Roediger, 2007; Pyc & Rawson, 2009; for a meta-analysis, see Rowland, 2014) and retrieval attempt interval—the time between studying and retrieving—increased (Pashler et al., 2003). Consistent with the present study, some studies have demonstrated the lack of a backward testing effect in inductive learning when participants were instructed to take an immediate test (e.g., Lee & Ahn, 2018; Lee & Ha, 2019).

Conclusion and Implications

As an increasing number of studies investigate JOL reactivity, it is critical to understand when the act of making JOLs aides or impedes learning. Previous research, however, predominantly focused on examining the impact of JOLs on verbal learning, primarily using retention tests to assess learning outcomes. To the best of our knowledge, this study is the first attempt to examine JOL reactivity in inductive learning while manipulating the extent to which JOLs are framed to elicit retrieval. Our results revealed that making typical forms of JOLs alone did not improve inductive learning; significant improvements were observed only when JOLs were accompanied by retrieval practice.

While the absence of JOL reactivity in inductive learning may not appear particularly intriguing, we believe that even a null difference offers valuable insights into the circumstances and mechanisms underlying JOL reactivity. From a theoretical perspective, our research suggests that JOL reactivity may change depending on the type of learning material and/or learning goal. Although many prior studies have reported positive JOL reactivity in simple verbal learning (Double et al., 2018), the null effect of making standard JOLs reported in text comprehension (Ariel et al., 2021; Zhao et al., 2023) and inductive learning from the current study suggests that more research is needed on how JOL reactivity varies depending on the complexity of the learning material. Furthermore, to gauge inductive learning, our study used a transfer test, where learners had to generalize what they learned from specific cases into other cases. This approach contrasts with most prior research, which has

focused on recall of information, primarily using retention tests (e.g., Ariel et al., 2021; Senkova & Otani, 2021). Students should be able to use appropriate levels of metacognitive judgments (Ha & Lee, 2023) and different learning strategies (Finley & Benjamin, 2012) to optimize their learning. Covert retrieval elicited while making JOLs may work as an effective strategy to enhance the retention of simple materials; however, it may be ineffective in promoting the comprehension and/or generalization of learning.

The present study also explored the possibility of using JOL as a learning strategy. Many previous studies reporting positive JOL reactivity suggest that the simple act of making JOLs can enhance memory. However, this study suggests that the use of JOL as a learning tool may not always be the best choice, especially considering our findings revealed that retrieval practice alone was as effective as and more time-efficient than the combined strategy of retrieval practice and JOL. With a large body of research on the testing effect (Roediger & Karpicke, 2006), this study again showed that retrieval practice is a robust learning tool. Although an increasing number of studies have investigated effective strategies for category learning (for a review, see Nosofsky & McDaniel, 2019), prior research has dominantly focused on model or theory development (Murphy, 2002). Our study contributes further evidence to the effectiveness of retrieval as a learning strategy in inductive learning, which has important implications for teaching practice and the development of study strategies.

From an educational perspective, understanding why the impact of making JOLs is limited in inductive learning can guide the development of more effective teaching strategies and learning tools tailored to the complex nature of inductive learning. This study suggests that students experience greater learning benefits when active retrieval is attempted in tandem with metacognitive judgments. To enhance student learning, educators may want to use retrieval practice alone as a study strategy or combine it with other metacognitive activities. For example, when teachers ask students to monitor and evaluate their own learning, they can encourage them to first retrieve the materials they have studied, and then make judgments of their learning status. Providing students with chances to be fully involved in retrieval practice may be a crucial step in enhancing the learning of new materials and should be undertaken before making metacognitive judgments.

Acknowledgment The present study is based on the first author's doctoral dissertation submitted to Yonsei University.

Preliminary results were presented as a poster at the meeting of the Cognitive Science Society, 2022. This study was not preregistered.

Funding This work was supported by the Yonsei University Research Grant of 2022 (2022-22-0207).

Declarations

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Approval was obtained from the Institutional Review Board of Yonsei University.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

- Amlund, J. T., Kardash, C. A. M., & Kulhavy, R. W. (1986). Repetitive reading and recall of expository text. *Reading Research Quarterly*, 49–58. <https://doi.org/10.2307/747959>
- Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, 138(3), 432–447. <https://doi.org/10.1037/a0015928>
- Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. K. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review*, 33, 693–712. <https://doi.org/10.1007/s10648-020-09556-8>
- Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, 41, 392–402. <https://doi.org/10.3758/s13421-012-0272-7>
- Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, 21(5), 279–283. <https://doi.org/10.1177/0963721412452728>
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, 34, 268–276. <https://doi.org/10.3758/BF03193405>
- Carvalho, M., Cooper, A., & Marmurek, H. H. (2023). Covert retrieval yields a forward testing effect across levels of successive list similarity. *Metacognition and Learning*, 18, 847–861. <https://doi.org/10.1007/s11409-023-09348-9>
- Chan, J. C., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018a). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective. *Journal of Memory and Language*, 102, 83–96. <https://doi.org/10.1016/j.jml.2018.05.007>
- Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018b). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, 144(11), 1111–1146. <https://doi.org/10.1037/bul0000166>
- Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *Quarterly Journal of Experimental Psychology*, 70(7), 1211–1235. <https://doi.org/10.1080/17470218.2016.1175485>
- Choi, H., & Lee, H. S. (2020). Knowing is not half the battle: The role of actual test experience in the forward testing effect. *Educational Psychology Review*, 32(3), 765–789. <https://doi.org/10.1007/s10648-020-09518-0>
- Dobson, J. L., Linderholm, T., & Stroud, L. (2019). Retrieval practice and judgements of learning enhance transfer of physiology information. *Advances in Health Sciences Education*, 24(3), 525–537. <https://doi.org/10.1007/s10459-019-09881-w>
- Double, K. S., & Birney, D. P. (2019). Reactivity to measures of metacognition. *Frontiers in Psychology*, 10, 2755. <https://doi.org/10.3389/fpsyg.2019.02755>
- Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, 26(6), 741–750. <https://doi.org/10.1080/09658211.2017.1404111>
- Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition*, 33(6), 1096–1115. <https://doi.org/10.3758/BF03193216>
- 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>
- Fiacconi, C. M., Mitton, E. E., Laursen, S. J., & Skinner, J. (2020). Isolating the contribution of perceptual fluency to judgments of learning (JOLs): Evidence for reactivity in measuring the influence of fluency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(5), 926–944. <https://doi.org/10.1037/xlm0000766>
- Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 632–652. <https://doi.org/10.1037/a0026215>
- Ha, H., & Lee, H. S. (2023). Think higher, gain more: The effect of making inference-and memory-based metacognitive judgments on text learning. *Metacognition and Learning*, 18(2), 567–590. <https://doi.org/10.1007/s11409-023-09341-2>

- Halamish, V., & Undorf, M. (2022). Why do judgments of learning modify memory? Evidence from identical pairs and relatedness judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(4), 547–556. <https://doi.org/10.1037/xlm0001174>
- Hartwig, M. K., & Dunlosky, J. (2017). Category learning judgments in the classroom: Can students judge how well they know course topics?. *Contemporary Educational Psychology*, 49, 80–90. <https://doi.org/10.1016/j.cedpsych.2016.12.002>
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1989). *Introduction: Processes of inference, learning, and discovery*. MIT Press.
- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1441–1451. <https://doi.org/10.1037/a0020636>
- Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review*, 25(6), 2356–2364. <https://doi.org/10.3758/s13423-018-1463-4>
- JASP Team. (2024). JASP (Version 0.18.3) [Computer software].
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Jönsson, F. U., Hedner, M., & Olsson, M. J. (2012). The testing effect as a function of explicit testing instructions and judgments of learning. *Experimental Psychology*, 59(5), 251–257. <https://doi.org/10.1027/1618-3169/a000150>
- Kang, S. H., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4–5), 528–558. <https://doi.org/10.1080/09541440601056620>
- Kang, Y., Ha, H., & Lee, H. S. (2023). When more is not better: Effects of interim testing and feature highlighting in natural category learning. *Educational Psychology Review*, 35(2), 51. <https://doi.org/10.1007/s10648-023-09772-y>
- Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151–162. <https://doi.org/10.1016/j.jml.2006.09.004>
- Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition*, 31(6), 918–929. <https://doi.org/10.3758/BF03196445>
- Kubik, V., Koslowski, K., Schubert, T., & Aslan, A. (2022). Metacognitive judgments can potentiate new learning: The role of covert retrieval. *Metacognition and Learning*, 17, 1057–1077. <https://doi.org/10.1007/s11409-022-09307-w>
- Lee, H. S., & Ahn, D. (2018). Testing prepares students to learn better: The forward effect of testing in category learning. *Journal of Educational Psychology*, 110(2), 203–217. <https://doi.org/10.1037/edu0000211>
- Lee, H. S., & Ha, H. (2019). Metacognitive judgments of prior material facilitate the learning of new material: The forward effect of metacognitive judgments in inductive learning. *Journal of Educational Psychology*, 111(7), 1189–1201. <https://doi.org/10.1037/edu0000339>
- Li, B., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., ..., Luo, L. (2021). Soliciting judgments of forgetting reactively enhances memory as well as making judgments of learning: Empirical and meta-analytic tests. *Memory & Cognition*, 50, 1061–1077. <https://doi.org/10.3758/s13421-021-01258-y>
- Li, B., Zhao, W., Shi, A., Zhong, Y., Hu, X., Liu, M., ..., Yang, C. (2023). Does the reactivity effect of judgments of learning transfer to learning of new information? *Memory*, 31(7), 918–930. <https://doi.org/10.1080/09658211.2023.2208792>
- Maxwell, N. P., & Huff, M. J. (2022). Reactivity from judgments of learning is not only due to memory forecasting: Evidence from associative memory and frequency judgments. *Metacognition and Learning*, 17(2), 589–625. <https://doi.org/10.1007/s11409-022-09301-2>
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15(1), 174–179. <https://doi.org/10.3758/PBR.15.1.174>
- Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, 145(2), 200–219. <https://doi.org/10.1037/a0039923>
- Miyatsu, T., Gouravajhala, R., Nosofofsky, R. M., & McDaniel, M. A. (2019). Feature highlighting enhances learning of a complex natural-science category. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 1–16. <https://doi.org/10.1037/xlm0000538>
- Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory*, 24(2), 257–271. <https://doi.org/10.1080/09658211.2014.1001992>

- Murphy, G. L. (2002). *The big book of concepts*. The MIT Press. <https://doi.org/10.7551/mitpress/1602.001.0001>
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of learning & motivation* (Vol. 26, pp. 125–173). Academic Press.
- Nosofsky, R. M., & McDaniel, M. A. (2019). Recommendations from cognitive psychology for enhancing the teaching of natural-science categories. *Policy Insights from the Behavioral and Brain Sciences*, 6(1), 21–28. <https://doi.org/10.1177/2372732218814861>
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1051–1057. <https://doi.org/10.1037/0278-7393.29.6.1051>
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 3–8. <https://doi.org/10.1037/0278-7393.31.1.3>
- Pastötter, B., & Bäuml, K. H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, 5, 286. <https://doi.org/10.3389/fpsyg.2014.00286>
- Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, 143(2), 644–667. <https://doi.org/10.1037/a0033194>
- Putnam, A. L., & Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition*, 41, 36–48. <https://doi.org/10.3758/s13421-012-0245-x>
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437–447. <https://doi.org/10.1016/j.jml.2009.01.004>
- Pyc, M. A., & Rawson, K. A. (2012). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 737–746. <https://doi.org/10.1037/a0026166>
- 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>
- 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>
- Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review*, 11(1), 131–136. <https://doi.org/10.3758/BF03206472>
- Schäfer, F., & Undorf, M. (2024). On the educational relevance of immediate judgment of learning reactivity: No effects of predicting one's memory for general knowledge facts. *Journal of Applied Research in Memory and Cognition*, 13(1), 113–123. <https://doi.org/10.1037/mac0000113>
- Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition*, 49(5), 955–967. <https://doi.org/10.3758/s13421-020-01133-2>
- Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 553–558. <https://doi.org/10.1037/a0038388>
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(1), 204–221. <https://doi.org/10.1037/0278-7393.26.1.204>
- Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory & Cognition*, 33(6), 1116–1129. <https://doi.org/10.3758/BF03193217>
- Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, 3(5), 315–317. <https://doi.org/10.1111/j.1467-9280.1992.tb00680.x>
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>

- Szpunar, K. K., Moulton, S. T., & Schacter, D. L. (2013). Mind wandering and education: From the classroom to online learning. *Frontiers in Psychology*, 4, 1–7. <https://doi.org/10.3389/fpsyg.2013.00495>
- Tauber, S. K., Dunlosky, J., & Rawson, K. A. (2015). The influence of retrieval practice versus delayed judgments of learning on memory: Resolving a memory-metamemory paradox. *Experimental Psychology*, 62(4), 254–263. <https://doi.org/10.1027/1618-3169/a000296>
- Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning? *Psychology and Aging*, 34(6), 1–12. <https://doi.org/10.1037/pag0000376>
- Tekin, E., & Roediger, H. L. (2021). The effect of delayed judgments of learning on retention. *Metacognition and Learning*, 16(2), 407–429. <https://doi.org/10.1007/s11409-021-09260-0>
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum.
- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, 18, 1140–1147. <https://doi.org/10.3758/s13423-011-0140-7>
- Witherby, A. E., & Tauber, S. K. (2017). The influence of judgments of learning on long-term learning and short-term performance. *Journal of Applied Research in Memory and Cognition*, 6(4), 496–503. <https://doi.org/10.1016/j.jarmac.2017.08.004>
- Witherby, A. E., Babineau, A. L., & Tauber, S. K. (2023). Does interactive imagery influence the reactive effect of judgments of learning on memory? *Journal of Intelligence*, 11(7), 1–12. <https://doi.org/10.3390/jintelligence11070139>
- Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied*, 23(3), 263–277. <https://doi.org/10.1037/xap0000122>
- Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(3), 485–492. <https://doi.org/10.1037/xlm0000449>
- Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *NPJ Science of Learning*, 3, 1–9. <https://doi.org/10.1038/s41539-018-0024-y>
- Zhao, W., Li, B., Shanks, D. R., Zhao, W., Zheng, J., Hu, X., ..., Yang, C. (2022). When judging what you know changes what you really know: Soliciting metamemory judgments reactively enhances children's learning. *Child Development*, 93(2), 405–417. <https://doi.org/10.1111/cdev.13689>
- Zhao, W., Xu, M., Xu, C., Li, B., Hu, X., Yang, C., & Luo, L. (2023). Judgments of learning following retrieval practice produce minimal reactivity effect on learning of education-related materials. *Journal of Intelligence*, 11(10), 190. <https://doi.org/10.3390/jintelligence11100190>

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.