한국심리학회지 : 인지 및 생물
The Korean Journal of Cognitive and Biological Psychology
2020, Vol. 32, No. 2, 169-188

Original Article
http://dx.doi.org/10.22172/cogbio.2020.32.2.004

Why They Chose What They Chose: Exploring Effects of Test Performance and Metacognitive Judgments on Learners' Selection of Interleaving Schedule*

Lan Anh Do[1], Hee Seung Lee[1†]

[1]Department of Education, Yonsei University

Effective learning demands awareness and spontaneous execution of an optimal study schedule. While prior research suggests the opposite, learners tend to believe that learning from a blocked schedule (grouping exemplars by category) is more effective than learning from an interleaved schedule (intermixing exemplars). We investigated how the initial learning experience affected learners' selection of a subsequent study schedule using a painting-style learning task. Participants studied the first section in an interleaved schedule and were given a different interim activity on that section (test vs. restudy). They were then asked to select their own study schedule for the second section between interleaved and blocked schedules. After that, participants took a final transfer test and again selected a subsequent study schedule. Participants were also asked to make judgments of learning (JOL) several times (after study and after test) throughout the experiment. The results revealed that the interleaving-selectors showed better learning on the subsequent section, demonstrating a robust interleaving effect. More interestingly, the first selection of interleaving schedule was predicted by learners' better interim-test performance. Furthermore, participants who experienced improvement via testing or via JOL throughout the experiment were more likely to stay on the same study schedule. These findings suggest that providing learners with a successful learning experience may be one way of encouraging them to choose and continue a more effective interleaving study method.

Keywords: effective study strategies, interleaving effect, study schedule selection, testing

1 차원고접수 19.12.11; 수정본접수: 19.03.19; 최종게재결정 19.03.26

How to improve students' learning behavior is a topic of growing interest due to the sizable body of research reporting on the frequent use of ineffective learning techniques (e.g., McCabe, 2011; Morehead, Rhodes, & DeLozier, 2016). Many students do not seem to be aware of empirically supported (ES) study strategies (McCabe, 2011). This is partly because effective learning strategies are often counter-intuitive; they may cause difficulties to the learning process and perhaps impede the performance at hand, although they eventually benefit long-term learning (referred as desirable difficulties, Bjork, 1994; Bjork & Bjork, 2011). Students often misinterpret such difficulties as unsuccessful learning, and fluency as actual learning progress, thus placing themselves at risk of engaging in easier but ineffective learning activities that generate the feeling of fluency

* This research was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A8020957), and in part by the Yonsei University Future-leading Research Initiative of 2017 (2017-22-0091).
† 교신저자: Hee Seung Lee, Department of Education, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Korea.
E-mail: hslee00@yonsei.ac.kr

(Bjork, Dunlosky, & Kornell, 2013). The problem is further exacerbated because the unawareness of ES study strategies appears to propagate among instructors (Morehead et al., 2015) and academic support centers (McCabe, 2018), who are supposed to serve as sources of knowledge about effective learning methods. Such a regrettable situation raises the need for more research on the factors that affect students' selection of study strategies, thus engaging more learners in effective study practice.

One example of the conflict between empirical evidence and students' beliefs is the benefit of interleaving (Kornell & Bjork, 2008; Yan, Bjork, & Bjork, 2016). A considerable body of research has shown that intermixing exemplars of different categories, which represent an interleaved schedule, is more effective than grouping the exemplars by category, a blocked schedule, in promoting long-term memory and induction (termed the interleaving effect; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Yan, Soderstrom, Seneviratna, Bjork, & Bjork, 2017). For example, when students study painting styles of various artists, the paintings of different artists may be presented in two different schedules. One is to intermix the paintings of different artists (interleaving), and the other is to group the paintings by artist (blocking). Kornell and Bjork (2008) had participants study a series of paintings in either blocked or interleaved order, and in a later test phase asked them to identify the artists of new paintings by the previously studied artists. Participants showed higher performance when they studied under the interleaved schedule than the blocked schedule, supplying convincing evidence for the superior effect of interleaving. However, many of the participants rated blocking as more effective than interleaving.

The interleaving effect has been reported in several fields of study, including mathematics (Patel, Liu, & Koedinger, 2016; Rohrer, Dedrick, & Stershic, 2015; Rohrer & Taylor, 2007), painting styles (Guzman-Munoz, 2017; Kang & Pashler, 2012; Kornell & Bjork, 2008; Yan et al., 2017), and bird families (Birnbaum, Kornell, Bjork, & Bjork, 2013; Kornell, Castel, Eich, & Bjork, 2010). There are two widely accepted explanations of the interleaving effect. First, the spacing-based account of the interleaving effect claims that when exemplars of different categories are interleaved, a time interval is formed between the first and second exposure to the exemplars of the same category. A temporal distribution among exemplars of the same category facilitates productive forgetting that can promote long-term retention and consequently enhance induction (Vlach & Kalish, 2014). Second, interleaving exemplars of different categories focuses learners' attention on between-category comparisons, thereby elucidating the important differences that distinguish one category from another, called the discriminative-contrast hypothesis (Birnbaum et al., 2013; Kang & Pashler, 2012).

Despite the robust effect of interleaving, few students are aware of the superiority of interleaved practice over blocking (Kornell & Bjork, 2008; Kornell et al., 2010; Yan et al., 2017; Zulkiply, McLean, Burt, & Bath, 2012). McCabe (2011) investigated students' prior beliefs on ES learning strategies employing six different pairs of scenarios that involved two contrasting study methods (i.e., one is empirically supported as being more effective than the other). Students were told to predict which situations would result in better learning outcomes. The findings showed that among the six learning strategies scientifically proven to be effective, the interleaved study was reported to receive the least endorsement, from less than ten percent of the participants. Recently, Anthenien, DeLozier, Neighbors, and Rhodes (2018) also showed that among ten different study strategies (from low to high effectiveness), students reported the least use of interleaving, even though interleaving was one of the most effective study strategies on the list. Besides prior belief, when given opportunities to schedule their learning, learners also showed a clear tendency to block, instead of interleaving exemplars of different categories (e.g., Carvalho, Braithwaite, de Leeuw, Motz, & Goldstone, 2016; Tauber, Dunlosky, Rawson, Wahleim, 2012; Yan et al., 2016). In the face of this undesirable phenomenon, it is important to understand the reasoning behind learners' decision-making and thus figure out under which circumstances we can engage more learners in the more effective interleaved practice.

How do learners gain knowledge about study strategy effectiveness?

Earlier studies suggest that learners can figure out whether a study strategy is effective or not by reflecting on their direct learning experience with that strategy (e.g., Brown, Smiley, & Lawton, 1978; Flavell, 1981). Such reflection can occur while learners take a test on previously learned content (i.e., an objective cue), and/or make a metacognitive judgment on their learning (i.e., a subjective cue). A large body of research, therefore, has utilized study-test experiences to help learners recognize the effectiveness of ES learning methods (e.g., Bjork, deWinstanley, & Storm, 2007; Brigham & Pressley, 1988; Burnett & Bodner, 2013; deWinstanley & Bjork, 2004; Pressley & Ghatala, 1989; Pressley, Levin, & Ghatala, 1984, 1988). Brigham and Pressley (1988), for example, provided empirical data showing the effect of a study-test experience on learners' awareness. They had participants who were native English speakers study a list of rare English words with a more effective keyword-mnemonic strategy (i.e., to make up a sentence using both the given keyword and the definition of the target word) and a less effective semantic-context strategy (i.e., to generate a sentence using the target word) alternatively. The participants later took a test on the previously learned word list. More importantly, they were asked to choose the more effective learning method at two timings. The first choice was made before the study and the second after the test. The results showed that, before the study, there was no significant preference for a particular strategy. However, young adult participants showed an apparent tendency to choose the more effective keyword-mnemonic strategy after the test.

Similarly, Pressley et al. (1984) also demonstrated that learners gain knowledge of the effectiveness of study strategies through their own study-test experiences in foreign vocabulary learning. They examined whether a study-test opportunity could help learners acknowledge the benefit of the elaboration strategy (i.e., to make a sentence linking the target word with a given keyword) over the repetition strategy (i.e., to repeat the target word over and over). Although the elaboration strategy was proved to be more effective, depending on the condition, participants were told that either the repetition strategy or the elaboration strategy was more effective. Then, half of the participants learned a word list by using the two different learning techniques alternatively and then took a test on their meanings (the practice condition). The other half of the participants were given only explanations about the two study methods but did not have a direct study-test experience (the no-practice condition). All the participants then decided on the study strategy for their subsequent study. Results showed that the participants in the practice condition reported a higher rate of selecting the more effective elaboration strategy regardless of the prior message that they received. This finding indicates the powerful effect of learners' study-test experience and suggests a pathway for gaining knowledge about the effectiveness of a study strategy.

The benefit of testing on strategy monitoring is also revealed in the forward effect of testing, which suggests that taking an interim test on previously learned content can improve the learning of new content over restudying it (Pastötter & Bäuml, 2014). For example, Lee and Ahn (2018) demonstrated the forward testing effect in the field of category learning using a painting-style learning task. In their study (Experiments 3 and 4), participants learned the painting styles of 12 different artists in an interleaved schedule across two sections, 6 artists in each section. After studying the first section, they either took an interim test (the interim-test condition) or restudied the paintings of the first section (the interim-restudy condition) before moving on to the second section. Participants then studied the painting styles of new artists in the same manner under both conditions. Although both groups of participants had identical study experiences with the second section, when they were asked to identify the corresponding artists of new paintings in a final transfer test, the interim-test group outperformed the interim-restudy group. The testing experience seems to allow learners to monitor the effectiveness of their current study strategies. Thus, when given an opportunity to study subsequently presented new material, they may implement more effective study strategies that in turn show enhanced performance in the subsequent learning phase (for a review, see Yang, Potts, & Shanks, 2018).

Besides testing, another way of evaluating learning strategies is for learners to make metacognitive judgments on their own learning. Several previous studies have shown that metacognitive assessment in various forms can influence learners' subsequent learning behavior, including the selection of restudy items (Kornell & Metcalfe, 2006; Morehead, Dunlosky, & Foster, 2017) and allocation of study time (Mitchum, Kelley, & Fox, 2016). More importantly, metacognitive monitoring can alter students' study strategy (e.g., Ben-Eliyahu & Bernacki, 2015; Pressley et al, 1984; Sahakyan, Delaney, & Kelley, 2004). For example, Sahakyan et al. (2004) found that, in vocabulary learning, asking learners to make metamemory judgment could help them adopt more effective encoding strategies for subsequent study. Ben-Eliyahu and Bernacki (2015) also suggested that having learners make metacognitive judgments can help them select an appropriate control strategy (e.g., note-taking) to elaborate their learning.

With that in mind, the present study aimed to examine the effect of both testing and metacognitive judgment on learners' selection of study schedule.

How might taking a test affect learners' selection of study schedule?

The answer varies depending on our approach to this question. Given the strong preference of learners toward blocking, because of both their prior belief and the encoding fluency induced by the blocked schedule, a single study-test experience might not be enough to convince the majority of learners of the benefits of the interleaving schedule (e.g., Kornell & Bjork, 2008; Kornell et al., 2010). For example, in Kornell and Bjork (2008, Experiments 1A and 2), almost 80% of the participants chose blocking as being more effective than interleaving even after they had actually experienced the superior effect of interleaving as reflected on their final test performance, implying that simply providing learning with experience may not be sufficient to correct learners' prior misbelief. To the best of our knowledge, Yan et al. (2016) was the only study that found a successful way of debiasing learners about the interleaving effect, which involved using both experience- and theory-based methods (Experiment 6). In Experiment 6, participants were explained the reasons why interleaving is more effective, why students often believe in the opposite, and that 90% of individuals show better performance with interleaved practice than with blocked study (the theory-based approach). Furthermore, participants encountered interleaved and blocked studying in two separate study-test cycles (the experience-based approach), with their experiences with interleaving starting first. Afterwards, more than 90 percent of the participants finally selected interleaving as more effective. The number dropped to 54% for those who received the same experience-based remedy but who were not given any theory. This finding again seems to suggest the limited effect of the study-test experience on learners' awareness of study schedule effectiveness.

However, what has been ignored in the aforementioned studies is the possibility that learners' selection of study schedule might depend on their personal experience: that is, their own test performance and judgments of learning. In other words, the success and failure of their previous learning experience can affect their subsequent study behavior. Bjorklund and Buchanan (1989) showed that the maintenance of a strategy use changed depending on the success or failure of earlier experience with that strategy. They had participants (i.e., school children from third to seventh graders) apply a categorization strategy to study several typical and atypical lists of exemplars across four trials. For the typical list, participants were better at categorizing the instances, and in turn were more likely to keep using the same categorization strategy across the four trials. The result was later replicated with college students using the same strategy (i.e., the categorization strategy; Rabinowitz, Freeman, & Cohen, 1992). Other studies expanded these findings by manipulating the degree to which related knowledge could be accessed and investigating how it affected the maintenance of strategy use. Similarly, they found that the more accessible the relevant knowledge was, the more likely students were to continue using a given strategy (see Rabinowitz & McAuley (2014) for the link mnemonic strategy; Woolley, Huang, & Rabinowitz (2019) for the componential strategy). These findings altogether suggest that one possible way of encouraging learners to choose a more effective study strategy is by providing them with successful learning experience with that strategy. For example, if students experienced a relative success with an interleaved schedule, then they might be more likely to continue interleaving in their subsequent study. For that reason, the current study not only considered the influence of the mere act of taking a test and making a metacognitive judgment but also examined how learners' test performance and their confidence in their learning affected the selection of their subsequent study schedule. Based on the findings of the previous studies, we predicted that the participants who experienced successful learning via testing and metacognitive judgment would be more likely to choose the more effective, interleaved schedule.

The present study

The present study has two main goals. The first was to examine how learners' selection of study schedule varied according to the presence of testing experience and their test performance. If testing plays an important role in evaluating the effectiveness of learners' current study strategy, then the presence of testing may help learners choose a more effective study method in their subsequent learning. Also, test performance may provide learners with a cue on how effective the current study method is. Therefore, consistent with previous studies, a successful learning experience that is reflected in higher test performance with the more effective interleaved schedule may lead learners to continue interleaving in their subsequent study. The second goal of the present study is to examine how metacognitive judgment affects learners' study-schedule choice. Similar to the effect of test performance, people who feel more confident (i.e., making a higher metacognitive judgment) about their initial learning in an interleaved schedule may evaluate their current schedule as being effective and would be more likely to choose interleaving when given another study opportunity.

To achieve the aforementioned goals, we used an interim-test effect test paradigm following the procedure of Lee and her colleagues (Lee & Ahn, 2018; Lee & Ha, 2019). Figure 1 illustrates the overall procedure of the present study. We had participants learn the painting styles of various artists across two sections (Section A and B) and administered different interim activities (test or restudy) between the two sections. All of the participants made a judgment of learning (JOL) three times across the experiment by predicting their performance on the artists of the corresponding section. Specifically, participants were asked to provide a JOL: right after studying Section A (JOL(A1)), after the interim test on Section A (JOL(A2)), and after Section B (JOL(B)). More importantly, participants were asked to choose their study schedule between interleaving and blocking for their subsequent study. In this way, we could examine how learners' initial learning experience affected the selection of their subsequent study schedule. Specifically, we looked at the impact of three factors, type of interim activity, learners' interim-test performance, and metacognitive judgments on their initial learning. Earlier studies only observed learners' metacognitive awareness of study schedules (e.g., Kornell & Bjork, 2008; Yan et al., 2016), but did not ask people to choose their own study schedule. However, what people endorse as effective is not always the same as what they actually practice (e.g., Hartwig & Dunlosky, 2012). Given the possible gap between belief and practice, it is important to investigate how learners spontaneously choose to use a study schedule when given the opportunity. However, in the design of the present study, the study schedule on Section B was not manipulated between subjects but was decided by participants, thus we cannot conclude the causal relationship between the study schedule and the final test performance. Nonetheless, the present study will help us understand how students' study strategies are related to their learning pefromance.

Method

Participants

A total of 61 undergraduate students (34 women, 27 men; mean age = 22.98 years) from a large university participated in exchange for a gift certificate equivalent to $5. However, one student in the interim–test condition was eliminated from the data analyses because of unexpected interruptions during the experiment, which resulted in a final sample size of 60 students.

Design

A one–way between–participants design was employed. The experiment manipulated the type of interim activity (test vs. restudy). Participants were either tested on the materials of Section A (interim–test condition) or restudied them (interim–restudy condition) before moving on to study new materials of Section B. Sixty participants were randomly assigned to one of two conditions, thus led to an equal sample size for the two conditions (n = 30 each).

Materials and procedure

The study was carried out in accordance with the Human Ethics Guidelines approved by the university where this research was conducted. We tested all the participants individually on a computer. The experiment started with an introduction regarding the purpose and general procedure of the study. The participants were informed that they would study the paintings of 12 artists over two sections and that there would be a test later on the studied content. Specifically, the participants were told that in the test phase, they would be presented with previously unseen paintings but were created by the artists whom they had studied. Their task would be to identify the artists corresponding to each of the new paintings. The present study used the same set of color landscapes paintings as in Lee and Ha (2019). These paintings were created by relative unknown artists and originally adapted from Kornell and Bjork (2008).[1) Among the 12 artists, the paintings of 6 artists were assigned to the first section (Section A), while those of the other 6 artists were assigned to the second section (Section B). We counterbalanced the artist–section pairs to control for specific item effect.

Figure 1 illustrates the basic schematic of the experimental procedures. In both conditions (interim–test and interim–restudy), all the participants started studying the paintings of Section A in an interleaved schedule. We started with the interleaved schedule because exposure to interleaved practice before blocked order (than the reversed order) is known to help more learners acknowledge the beneficial effect of interleaving (Yan et al., 2016). Furthermore, it is necessary to have all the participants successfully learn initial learning materials, before being asked to choose the more effective study schedule for the subsequent learning section. Thus, following the previous findings that the interleaved schedule promoted learning better than a blocked one (Kornell & Bjork, 2008; Kornell et al., 2010), we had all the participants start with the interleaved schedule. Participants studied the same set of 36 paintings (six paintings from each of six artists) in a fixed random order. Each of the paintings was presented simultaneously with the corresponding artist's name written below the painting for 5 seconds. Each painting was followed by a 0.5–second blank screen. After finishing the first study section, the participants were instructed to make a JOL by predicting their test performance on the artists of Section A. Specifically, they were asked what percentage (0 - 100) they would correctly answer if they were presented with previously unseen paintings created by the studied artists as in Section A and had to identify the

1) Lee and Ha (2019) did not include the paintings by Ciprian Stratulat that were used in Kornell and Bjork (2008) due to their low resolution but replaced them with paintings by Emma Ciadi (the original paintings were retrieved from https://sites.williams.edu/nk2/)

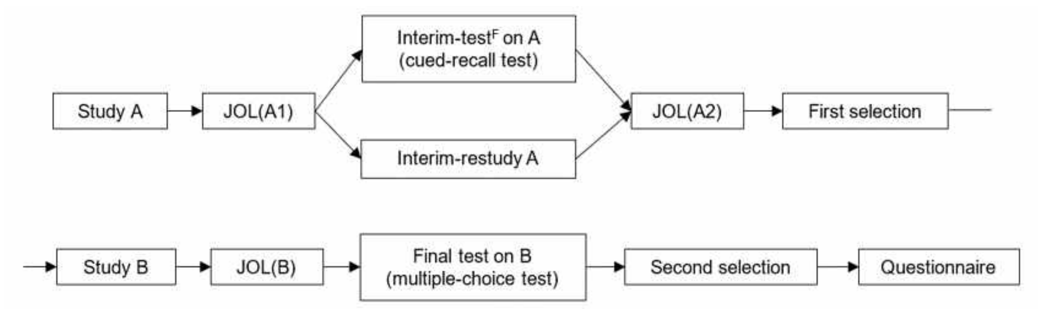

Figure 1. Schematic of the basic experimental procedure. Interim-test^F indicates interim test with feedback. JOL(A1) and JOL(A2) represent metacognitive judgments of Section A, and JOL(B) represents a metacognitive judgment of Section B.

painter of each. The participants made JOLs a total of three times throughout the experiment. Therefore, from now on we will call the JOL made immediately after the study session for Section A as JOL(A1) to discriminate it from the other two JOLs.

After making JOL(A1), the participants took an interim test (interim-test condition) or restudied a part of the studied materials (interim-restudy condition). During the interim activity, the same set of paintings was used in both conditions. They consisted of two paintings randomly selected from each of the six artists studied, for a total of 12 paintings. In the interim-restudy condition, the paintings were presented in the same manner as the study session. In the interim-test condition, the paintings were presented without the name of the artists. The participants were asked to type the name of the artist of each of the paintings (cued-recall format). There was no time limit, and feedback was provided for 2 seconds after the participants submitted their answers. The feedback page presented the painting simultaneously with the corresponding artist's name. We provided feedback to control for the exposure to the materials of Section A between the two conditions. The interim activity was followed by the second JOL on the artists of Section A, JOL(A2), under both conditions.

All the participants were told that they would move to the second section (Section B) and study the paintings by six artists different from those in Section A. However, this time they were asked to select the schedule in which the paintings would be displayed, between a blocked or interleaved schedule. The participants were encouraged to select the schedule that they thought would be more effective for their learning. Two options were provided like following. The first option is the same schedule as Section A—that is, the paintings would be randomly intermingled. The second option is that all the paintings by the same artist would be grouped together. Figure 2 illustrates how the paintings were distributed in the interleaved and blocked schedules.

In Section B, according to the participants' own selection, the participants studied a set of 36 paintings by a different group of six artists (six paintings from each of six artists) in either an interleaved or a blocked schedule. The paintings were presented in the same manner as in Section A except for the study schedule (only if the participants selected a blocked schedule for Section B). The study session was followed by the third JOL. However, this time the question concerned the artists of Section B, JOL(B). After making JOL(B), the participants moved immediately to the final test on the artists of Section B. The final test was a multiple-choice transfer test. We presented new paintings from previously

Interleaved	B1	D1	A1	F1	C1	E1
Blocked	A1	A2	A3	A4	A5	A6

Figure 2. Schematic of the interleaved and blocked schedule. The letters A, B, C, D, E, F represent different artists while the numbers next to the letters indicate the ordinal number of the painting of the respective artist.

learned artists and participants had to identify the corresponding artist for each painting from six different options. Because the format of the final test was different from the interim cued-recall test, no participant had been exposed to the format of the final test in advance and all participants were tested under the same conditions. There were 12 paintings on the final test (2 new paintings by each studied artist on Section B). There was no time limit and no feedback provided. Upon completion of the final test, the participants made the second selection of a study schedule for the third section (interleaved vs. blocked), even though no third section followed. Finally, the participants were told to fill out a post-experiment questionnaire on a separate sheet of paper. The questionnaire asked them what they had selected for the third section (the second selection) and the reasons for their selections. After that, the participants were debriefed and thanked.

Results and Discussion

The present study divided the participants by both condition (interim-test vs. interim-restudy) and the participants' first selection (interleaving vs. blocking), resulting in four groups: test-interleaving^{S1}, test-blocking^{S1}, restudy-interleaving^{S1}, and restudy-blocking^{S1}. The superscript S1 indicates the first selection of the study schedule. The test-interleaving and test-blocking groups represent the participants in the interim-test condition that chose interleaving and blocking at the first selection, respectively. The restudy-interleaving and restudy-blocking groups include the participants in the interim-restudy condition that chose interleaving and blocking at the first selection, respectively. Table 1 presents a data summary of these four groups.

Was the interleaving effect observed?

In Section B, the participants studied the paintings following the schedule of their own choice. Thus, under the same condition, if the participants who chose the interleaved schedule for Section B (the interleaving-selectors^{S1}) show a higher final transfer performance compared with those who chose the blocked schedule (the blocking-selectors^{S1}), we might interpret that the interleaving effect occurred.

A 2 (interim activity: interim-test vs. interim-restudy) x 2 (first selection of study schedule: interleaving vs. blocking) between-subjects analysis of variance (ANOVA) was conducted on the mean percentage of correct responses. There was no main effect of the interim activity, $F(1, 56) = 0.19$, $p = .669$, $\eta_p^2 = .003$. However, there was a significant main effect of the first selection, $F(1, 56) = 18.23$, $p < .001$, $\eta_p^2 = .246$, in that the interleaving-selectors^{S1} performed significantly better in the final transfer test ($M = 82.94$, $SD = 12.21$) than the blocking-selectors^{S1} ($M = 58.76$, $SD = 24.70$) regardless of the condition. There was no significant interaction between the interim activity and first selection, $F(1, 56) = 2.85$, $p = .097$, $\eta_p^2 = .048$.

To test whether the interleaving-selectors^{S1} outperformed the blocking-selectors^{S1} in the final test, two independent t-tests were conducted. The results showed that the test-interleaving^{S1} group performed significantly better on the final test than the test-blocking^{S1} group, $t(28) = 4.23$, $p < .001$, $d = 1.60$. Similarly, the restudy-interleaving^{S1} group also outperformed the restudy-blocking^{S1} group, $t(27) = 2.17$, $p = .040$, $d = 0.84$.[2] Overall, these results revealed the interleaving effect in inductive learning.

Why they chose what they chose: First selection

To examine whether participants' selection of study schedule at the first selection varied by condition, a chi-square test of independence was conducted on the selection ratios. There was no significant difference between the interim-test and interim-restudy conditions in the first selection ratios, $\chi^2(1) = 0.07$, $p = .787$, suggesting that testing did not necessarily encourage participants to choose a more effective study strategy. The selection ratios at the first selection (as shown in Table 1) were almost the same despite different interim activities administered in the two conditions. About

2) The degrees of freedom were adjusted when the Levene's test indicated unequal variances.

Table 1. Data summary based on the interim activity condition and first selection of study schedule. The table includes the mean JOL, mean accuracy of the interim and final test, and the percentage of interleaving selections.

			Interim test (test on A)							
Conditions	*N*	JOL(A1)	Block 1	Block 2	Average	JOL(A2)	First selection	JOL(B)	Final test (test on B)	Second selection
Interim-test	30	63.60 (18.68)	52.78 (23.20)	57.78 (25.79)	55.28 (21.94)	54.50 (26.08)	33% (10/30)	56.50 (19.88)	64.17 (25.91)	57% (17/30)
Test-interleaving^{S1}	10	68.50 (9.44)	61.67 (22.29)	78.33 (13.72)	70.00 (14.12)	67.50 (21.25)		66.50 (13.75)	86.70 (11.91)	70% (7/10)
Test-blocking^{S1}	20	61.15 (21.71)	48.33 (22.88)	47.50 (24.35)	47.92 (19.45)	48.00 (26.28)		51.50 (20.84)	52.80 (23.67)	50% (10/20)
Interim-restudy	30	60.87 (25.94)	-	-	-	69.77 (23.37)	37% (11/30)	67.43 (23.91)	70.28 (22.07)	40% (12/30)
Restudy-interleaving^{S1}	11	76.91 (23.11)	-	-	-	82.36 (21.76)		76.73 (25.32)	79.64 (12.10)	73% (8/11)
Restudy-blocking^{S1}	19	51.58 (23.22)	-	-	-	62.47 (21.54)		62.05 (21.54)	64.95 (24.85)	12% (4/19)

Note. S1 indicates the first selection. Test-interleaving and test-blocking represent the participants in the interim-test condition that chose interleaving or blocking at the first selection. Restudy-interleaving and restudy-blocking refer to the participants in the interim-restudy condition that chose interleaving or blocking at the first selection. Block 1 and Block 2 represent the mean percentages of correct responses in the former half and the latter half of the interim test. The numbers under the First Selection and the Second Selection columns show the percentages of participants who chose interleaving. The numbers in the parentheses under the First Selection and the Second Selection columns describe the proportions of participants who selected interleaving at the first and the second selection. All the numbers in the other parentheses represent the standard deviations.

two-thirds of the participants under each condition selected a blocked schedule for Section B, which reflected the robust preference of learners towards blocking. However, another important goal of the present study was to explore how test performance and metacognitive judgments influenced participants' selection of the study schedule. Thus, we examined whether there was a significant difference in the interim-test performance and in the JOL between the test-interleaving^{S1} and test-blocking^{S1} groups. In the interim-restudy condition, as participants did not take any test in Section A, we only compared the JOLs between the restudy-interleaving^{S1} and restudy-blocking^{S1} groups.

Interim-test condition. Because participants were given immediate feedback on every trial, we expected that their performance would get better in the second half of the test. Thus, we divided the interim-test into two blocks of six trials that we called Block 1 (i.e., the former six trials) and Block 2 (i.e., the latter six trials). To compare the performance of the test-interleaving^{S1} and test-blocking^{S1} groups, multiple independent t-tests were conducted on the mean percentage of correct responses. Regarding the average performance for both blocks, the test-interleaving^{S1} group outperformed the test-blocking^{S1} group, $t(28) = 2.92$, $p = .007$, $d = 1.10$. However, the performance patterns were more apparent in Block 2 than in Block 1. For Block 1, there was no significant difference between the two groups, $t(28) = 1.52$, $p = .140$. In contrast, for Block 2, the test-interleaving^{S1} group performed significantly better than the test-blocking^{S1} group, $t(27) = 4.43$, $p < .001$, $d = 1.71$, suggesting that high performing participants were more likely to choose interleaving over blocking schedule for their subsequent study.

In order to examine the effect of participants' JOLs on the selection of study schedule, two independent t-tests were performed on the mean ratings of JOL(A1) and JOL(A2) between the test-interleaving[S1] and test-blocking[S1] groups. For JOL(A1), there was no significant difference between the two groups, $t(28) = 1.29$, $p = .208$. However, the difference in the mean ratings of JOL(A2) between the test-interleaving[S1] and test-blocking[S1] groups was marginally significant, $t(28) = 2.03$, $p = .052$, $d = .77$. The test-interleaving[S1] reported numerically higher JOL(A2) than the test-blocking[S1] group (82.36 and 62.47, respectively), implying that participants who reported higher confidence on their learning were more likely to choose interleaving over a blocking schedule for their subsequent study.

We also conducted a logistic regression to determine the association between the interim-test performance, JOL(A1, A2), and the first selection of study schedule. The model entailed the selection of study schedule as a binary criterion variable, "1" for interleaved and "2" for blocked schedule, and four continuous predictor variables, JOL(A1), JOL(A2), and Block 1 and Block 2 performances (interim-test). The results are presented in Table 2. Among the four predictors, only the Block 2 performance of the interim-test significantly predicted participants' selection of study schedule at the first selection, $p = .022$, suggesting that those who performed relatively better in the second block of the interim-test on Section A were more likely to have selected an interleaved schedule for Section B.

Table 2. Logistic regression on the two timings of selection and on the direction of schedule selection from the first to the second selection in the two interim activity conditions.

Interim-test condition							
	First Selection			Final Selection		Direction of Selection	
Variables	B	Odd ratios	Variables	B	Odd ratios	B	Odd ratios
JOL(A1)	0.03	1.03	JOL(B)	-0.01	0.99	0.01	1.01
JOL(A2)	-0.04	0.96	Final test	-0.03	0.98	-0.01	0.99
Block 1 (interim-test)	0.05	1.05	Performance improvement	0.07*	1.08	-0.06*	0.94
Block 2 (interim-test)	-0.13*	0.88		-	-	-	-
Nagelkerke r^2	56.26%		Nagelkerke r^2	44.19%		40.80%	
χ^2	15.58		χ^2	12.06		10.88	
	$df = 4$, $p = .004$			$df = 3$, $p = .007$		$df = 3$, $p = .013$	
Interim-restudy condition							
	First Selection			Final Selection		Direction of Selection	
Variables	B	Odd ratios	Variables	B	Odd ratios	B	Odd ratios
JOL(A1)	-0.07	0.94	JOL(B)	-0.04	0.96	$< .01$	1
JOL(A2)	0.02	1.02	JOL improvement	0.04	1.041	-0.09*	0.91
	-	-	Final test	-0.01	0.99	0.02	1.02
Nagelkerke r^2	31.19%		Nagelkerke r^2	24.13%		42.20%	
χ^2	7.77		χ^2	5.90		10.26	
	$df = 2$, $p = .021$			$df = 3$, $p = .117$		$df = 3$, $p = .016$	

Note. Block 1 and Block 2 represent the mean percentages of correct responses in the former half and the latter half of the interim test. Performance improvement was calculated by subtracting the mean percentages of correct responses in the latter half of the interim test (Block 2) from those in the final test. JOL improvement was calculated by subtracting the mean ratings of the second metacognitive judgments of Section A (JOL(A2)) from those of the metacognitive judgment of Section B (JOL(B)).

*$p < .05$.

Interim-restudy condition. To investigate the effect of participants' JOLs on the selection of study schedule, two independent t-tests were conducted on the mean ratings of JOL(A1) and JOL(A2) between the restudy-interleaving^{S1} and restudy-blocking^{S1} groups. The results revealed that the restudy-interleaving^{S1} group reported a significantly higher JOL than the restudy-blocking^{S1} group at both JOL(A1), $t(28) = 2.88$, $p = .007$, $d = 1.09$, and JOL(A2), $t(28) = 2.43$, $p = .022$, $d = .09$, implying that people who were more confident about their learning in Section A were more likely to choose interleaving for Section B. As a consequence, we conducted a logistic regression to examine the relationship between participants' JOLs and their first selections in the interim-restudy condition. However, as shown in Table 2, JOL(A1) and JOL(A2) were not found to be significant predictors of participants' first selection.

In short, based on the analyses of the first selection, there was a tendency for those who showed a relatively more successful learning experience (better performance on the interim-test and higher metacognitive judgment) on Section A to choose the interleaved schedule for Section B. However, one should be careful in drawing any conclusions from this result. Selecting the interleaved schedule at the first selection could be interpreted in two different ways. One interpretation is that perhaps those who had relatively better performance on the interim test (the interim-test condition) and felt more confident about their learning (the interim-restudy condition) recognized the interleaving effect, and thus explicitly chose the interleaved schedule for their subsequent study. An alternative interpretation is that those who acquired initial learning success with the interleaved order decided to continue with the same study schedule that they had experienced in Section A, whereas those who performed relatively poorly and felt less confident decided to try a different study schedule. A closer look at the second selection would be needed to understand the reasoning behind the participants' decision.

Why they chose what they chose: Second selection

To examine whether participants' second selections varied by interim activity conditions, a chi-square test of independence was conducted on the selection ratios. There was again no significant difference between the two conditions in the second selection ratios, $\chi^2(1) = 1.67$, $p = .196$, implying that test experience itself did not directly affect learners' selection of study schedule. However, more positive results were found at the second selection than at the first. Specifically, under the interim-test condition, an additional 20 percent of the participants selected the interleaved schedule at the second selection (i.e., 57%) than at the first (i.e., 33%). On the contrary, in the interim-restudy condition, the selection ratios of the study schedule at the first (i.e., 37%) and second selections (i.e., 40%) were almost the same (see Table 1). Two McNemar tests conducted on the selection ratios, however, showed that there was no significant difference between the first and the second selections of the participants under the interim-test condition, $p = .092$, and under the interim-restudy condition, $p = 1.000$.

It is worth noting that even though all the participants under both conditions went through the same procedure on Section B, the final test was the only test that participants encountered across the experiment under the interim-restudy condition. On the other hand, under the interim-test condition, the final test was the second test that participants encountered, after the interim-test on Section A. Thus, as the final test performance might have played different roles in participants' second selection across the two conditions, it is necessary to examine the two conditions separately. Similar to the analysis of the first selection, we again divided the participants into four groups based on both the interim activity and participants' second selection, resulting in four groups of participants: the test-interleaving^{S2}, test-blocking^{S2}, restudy-interleaving^{S2}, and restudy-blocking^{S2} groups. The superscript S2 indicates the second selection. The test-interleaving and test-blocking groups represent the participants in the interim-test condition that chose interleaving and blocking at the second selection, respectively. The restudy-interleaving and restudy-blocking groups refer to the participants in the interim-restudy condition that chose interleaving and

Table 3. Data summary based on the interim activity condition and second selection of study schedule. The table includes the mean JOL, mean accuracy of Block 2 (the interim test) and the final test, and the percentage of interleaving selections at the first selection.

Conditions	*N*	Block 2 (test A)	JOL(A2)	First select	JOL(B)	JOL improve	Final test (test B)	Perf improve
Interim-test								
Test-interleaving^{S2}	17	68.63 (14.29)	67.94 (18.63)	52% (7/17)	58.53 (18.27)	−9.41 (19.99)	62.25 (27.34)	−6.37 (24.21)
Test-blocking^{S2}	13	43.59 (30.84)	36.92 (24.29)	19% (3/13)	53.85 (22.28)	16.92 (21.17)	66.67 (24.77)	23.08 (24.80)
Interim-restudy								
Restudy-interleaving^{S2}	12	_	81.33 (20.94)	67% (8/12)	75.00 (27.55)	−6.33 (15.15)	77.08 (15.94)	_
Restudy-blocking^{S2}	18	_	62.06 (22.15)	17% (3/18)	62.39 (20.42)	0.33 (23.79)	65.74 (24.73)	_

Note. S2 indicates the second selection. Test-interleaving and Test-blocking represent the participants in the interim-test condition that chose interleaving or blocking at the second selection. Restudy-interleaving and restudy-blocking refer to the participants in the interim-restudy condition that chose interleaving or blocking at the second selection. Block 2 represents the mean percentages of correct responses in the latter half of the interim test. JOL improvement was calculated by subtracting the mean ratings of the second metacognitive judgments of Section A (JOL(A2)) from those of the metacognitive judgment of Section B (JOL(B)). Performance improvement was calculated by subtracting the mean percentages of correct responses in the latter half of the interim test (Block 2) from those in the final test. The numbers in percentage under the First Selection column show the percentages of participants who chose interleaving. The numbers in the parentheses under the First Selection column describe the proportions of participants who selected interleaving at the first selection. All the numbers in the other parentheses represent the standard deviations.

blocking at the second selection, respectively. The overall results are shown in Table 3.

Interim-test condition. Because participants under the interim-test condition engaged in two tests, one on each section, a subjective feeling of how much better or worse they performed between the two tests might influence their second selection. We anticipated that the performance improvement from the interim test to the final test would also have an impact on participants' second selection. The improvement from the interim test to the final test was calculated by subtracting the mean percentages of correct responses on the latter half of the interim test (Block 2) from those on the final test. This is termed *performance improvement* in the present study.

We conducted a logistic regression to identify the association between participants' JOL(B), final test performance, performance improvement from Section A to B, and their second selection of study schedule. Table 2 shows the results of the logistic regression. Among the three variables included in the model, only the performance improvement significantly predicted participants' second selection, p = .014, suggesting that the larger improvement participants experienced, the more likely they were to choose the blocked schedule at the second selection. This result casts doubt on the first interpretation of the analysis for the first selection, which stated that relatively better performers on Section A would recognize the benefits of interleaving and thus select the interleaved schedule for Section B.

In order to test our second interpretation that better performers would continue to choose the same study schedule, we created a new variable called *direction of selection*. We coded participants "0" if they picked the same study schedule two times (same direction), and "1" if they ever changed their selection of study schedule (different direction). We conducted a logistic regression to investigate the relationship among the JOL(B), the final test performances, the performance improvement, and participants' direction of selection. Table 2 shows the results. As we expected, only the performance improvement significantly predicted participants' direction of selection, p = .043, implying that the more improvement participants experienced, the more likely they were to persist in the same study schedule as their previous learning experiences. This result is consistent with our second interpretation of the analysis of participants' first selection such that participants would choose the same study schedule if their performance improved from Section A to B. In contrast, they would be more likely to alter their study schedule if little or no improvement occurred.

Interim-restudy condition. Unlike the interim-test condition, the interim-restudy group did not take an interim-test and therefore we could not measure improvement in their test performance. Instead, we created a new variable, *JOL improvement*, which was calculated by subtracting the mean ratings of JOL(A2) from those of JOL(B).

First, consistent with our procedure for the interim-test condition, we conducted a logistic regression to examine the association between participants' JOL(B), JOL improvement, final test performance, and second selection. Table 2 shows the results of the regression analysis. None of the variables was found to significantly predict the participants' second selection.

Second, we examined the data divided by the two conditions and participants' direction of selection from the first selection to the second selection. As shown in Table 2, a logistic regression again revealed that the JOL improvement significantly predicted participants' direction of selection in the interim-restudy condition, p = .039. Altogether these results support our prediction that participants tend to consider their relative learning performance at different time points before deciding whether to continue with the same or a different study schedule.

Post-questionnaire

The current study asked participants to explicitly report why they chose what they chose at the second selection in a separate post-questionnaire at the end of the experiment. We discerned a frequent appearance of the words "compare," "difference," and "commonality." Of 60 responses, there were 18 participants (i.e., 30%) mentioning "compare," 9 participants (i.e., 10%) mentioning "difference," and 10 participants (i.e., 17%) mentioning "commonality" in their response. More interestingly, many participants who selected interleaving appeared to emphasize the importance of finding differences among the artists' styles whereas many participants who selected blocking tended to mention the importance of finding commonalities across paintings within an artist.

In order to examine this observation, we investigated whether each participant emphasized difference or commonality in his or her written response. Two research assistants who were blind to our research analyzed the participants' responses in the post-questionnaire. There was a substantial agreement between the two raters, κ = .655 (95% CI, .498 to .812), $p < .001$. When their opinions did not overlap, the final responses were decided by a third rater.

Table 4 describes the analysis of the participants' responses. Regardless of the conditions, most of the participants who selected interleaving at the second selection (i.e., 62%) indicated an emphasis on detecting the differences among the artists' styles. For example, one of the participants reported that (she selected interleaving because) it was easy to compare the differences between one artist and another when studying in an interleaved manner. However, in the case of blocked order, even though the characteristics of each artist were easy to spot, it was difficult for her to tell the differences between one artist and others that shared similar characteristics.[3] In contrast to those who selected interleaving, almost half of the participants who selected blocking at the second selection (i.e., 48%) stated the importance of finding the commonalities among different

3) Participants were not exposed to the word "interleaving" and "blocking" during the experiment. Therefore, in the response, she wrote "the first method," which implied interleaved order, and "the second method," which represented the blocked order.

Table 4. The percentage of participants who emphasized the importance of difference, commonality, both, or others in the post questionnaire.

Second Selection	Emphasis				Total
	Difference	Commonality	Both	Others	
Interleaving	62% (18/29)	0% (0/29)	0% (0/29)	38% (11/29)	100% (29/29)
Blocking	10% (3/31)	48% (15/31)	3% (1/31)	39% (12/31)	100% (31/31)

Note. The numbers in the parentheses represent the number of participants

paintings of one artist. For instance, one participant specified that (she chose blocked order because) grouping the paintings by artists made it easier for her to find the commonalities among the paintings of one artist. This pattern revealed one important reason behind participants' selection of study schedule. It appears that the goal regarded as key to mastering the inductive task led them to select a study schedule that seemed likely to facilitate their goal. Additional discussion of this finding in connection with previous studies will be provided in the General Discussion.

General Discussion

Students spend a substantial amount of time studying on their own; thus it is important to understand how they spontaneously select their study method and what factors affect this decision-making process. The present study was the first to reveal how learners' experience on their prior learning influenced their selection of a subsequent study schedule. Participants in the present study engaged in two study sections and could select subsequent study schedules two times (each time following one section). In this way we were able to examine the effect of test performance and JOL on the selection of a subsequent study schedule in a continuous timeline.

The results showed that under the interim-test condition, the performance in the latter half of the interim test significantly predicted the selection of study schedule for Section B, suggesting that participants used their test performance as a cue for deciding how to study next. The better the test performance was, the more likely participants chose the more effective interleaved schedule over the less effective blocked schedule. In contrast, under the interim-restudy condition, test performance was not available as a cue; rather, only subjective metacognitive judgments (i.e., JOL(A1, A2)) were available for deciding the subsequent study method. Indeed, there was a numerical tendency that those who felt more confident about their learning experience with the interleaved schedule decided to interleave their subsequent study. However, the metacognitive judgments were not found to be significant predictors of participants' first selection. In short, regardless of the interim activities, we found a consistent pattern that the people who selected interleaving for Section B were those with more successful learning (higher interim-test performance or higher metacognitive judgments) in Section A than those who selected blocking. These results could be interpreted in two ways. First, the high-performing learners on previous sections were better able to recognize the advantage of interleaved order, and thus selected interleaving for their subsequent study. Another interpretation was that those who succeeded with the interleaved presentation of paintings in Section A continued to use the same study method for Section B.

Further analysis of the second selection supported the latter explanation. Both the performance improvement and JOL improvement throughout the experiment significantly predicted participants' direction of schedule selection under the interim-test condition and the interim-restudy condition, respectively. Those who experienced improvement via testing or via JOL tended to continue with the same study schedule, whereas those who had little or no improvement were more likely to change their study schedule. It appears that the participants relied on the success and failure of their own study experience to assess the effectiveness of the study schedule. Thus, when their performance actually improved and/or they felt their performance improved, they probably thought that the current study schedule was effective, and therefore decided to continue using it.

We also examined the reasoning behind several participants' selections of their study schedule by analyzing the participants' responses on the questionnaire. Specifically, we found that 48% of those who chose blocking explained that blocking was more effective because it facilitated the finding of commonalities among the paintings of the same artist. They thought that success in the painting styles induction task relied on the identification of commonalities across the paintings that defined the style of one artist. In contrast, 62% of the participants who chose interleaving reported that interleaving helped them find differences among the artists' painting styles. Such a response is consistent with the explanations of the superiority of interleaved schedule provided in previous studies. Kang and Pashler (2012) argued that interleaving was more effective than blocking because an interleaved schedule allowed learners to compare and contrast various paintings of different artists, thus helping them to identify the critical differences among the artists' painting styles. The interleaving selectors appeared to be aware of what is critical for successful inductive learning. That is, because they thought that interleaving would better highlight the distinction between one artist and another, they would therefore be able to learn better with the interleaving schedule.

However, we did not find any benefit of testing in helping learners choose the more effective study method. At the first selection, the proportion selecting the interleaved order was almost the same under the two interim activity conditions (i.e., approximately 30%), indicating that our manipulation of the interim activity (test vs. restudy) did not influence the participants' selection of study schedule. The results differed at the second selection—that is, there were more participants under the interim-test condition (i.e., 57%) selecting interleaved schedule at the second selection than under the interim-restudy condition (i.e., 40%). The difference between the two conditions, though, was not significant. These findings are consistent with previous studies showing that a testing experience alone did not help the majority of learners to acknowledge the superior effect of interleaving (e.g., Kornell & Bjork, 2008; Kornell et al., 2010; Yan et al., 2016).

Why was there no benefit of the interim-test activity over the interim-restudy at the first selection? We propose three possibilities. First, this is perhaps due to the poor performance of most participants in the cued-recall interim test. In particular, those who selected blocking at the first selection (about two-thirds of the participants in the interim-test condition) reported average scores of 48.33% and 47.50% in the first and second blocks of the interim test, respectively. They correctly answered less than half of the total questions. Given that feedback was provided for every trial during the interim test, such scores represent quite poor performance and participants might have incorrectly attributed their low performance to the study schedule (i.e., interleaving), thus leading to the selection of a different order for subsequent study (i.e., blocking). This is very likely because, as participants had not experienced the blocked order, there was no baseline for comparison. Indeed, the tendency to choose the same or a different study schedule at the second selection differed according to the test performance. Second, because all the participants started with an interleaved schedule and did not experience the blocked schedule, they probably lacked the opportunity to compare the efficacies of the two schedules. Pressley et al. (1988) found that practice with both the effective and ineffective strategies led to higher maintenance of the effective strategy than did practice with the effective strategy alone. Pressley et al. (1984) also suggested that comparing and contrasting different strategies—referred to as Metamemory Acquisition Procedures (MAP)—can lead to substantial knowledge about strategy utility, the knowledge that students need when making decisions on their strategy use. Indeed, many studies that showed the facilitating effect of test taking on strategy-efficacy awareness allowed participants to practice both the effective and ineffective strategies (e.g., Brigham & Pressley, 1988; Pressley et al., 1984). Third, in earlier studies that demonstrated the benefit of testing on strategy awareness, participants did not seem to hold a strong bias toward the nonoptimal learning method (e.g., Brigham & Pressley, 1988; Pressley et al., 1984). For example, according to Brigham and Pressley (1988), prior to the study-test trial, participants judged the two to-be-practiced strategies (i.e., semantic-context and keyword-mnemonic) as likely to work similarly well. In the present study, however, blocking is more widely endorsed (e.g., McCabe, 2011; Morehead et al., 2016) and practiced (e.g., Carvalho et al., 2016; Tauber et al., 2012; Yan et al., 2017) by learners than is interleaving. This a priori belief posed a much more difficult challenge to the debiasing process.

Theoretical and Practical Implications

The present study provides important theoretical and practical implications. From a theoretical standpoint, consistent with earlier studies (e.g., Kang & Pashler, 2008; Kornell & Bjork, 2008), it replicates the highly robust effect of interleaving in inductive category learning. However, because the study schedule was not manipulated between subjects but was decided by the participants themselves, the present study cannot conclude on a causal relationship between the study schedule and final test performance. Furthermore, the present study expands our knowledge about the effects of testing on learners' study strategy monitoring. Many previous studies have suggested that a testing experience could improve learners' strategy-efficacy awareness (e.g., Lee & Ahn, 2018; Pressley et al., 1984). However, the present study shows that when the ES learning method counters learners' belief, a testing experience alone is not sufficient to encourage the majority of students to adopt the ES study strategy. Performance and JOL improvement after two times of testing can even reinforce the blocking behavior if learners might mistakenly attribute their improvement to the blocked order, thereby continuing to select blocking. However, one strength of the present study is that it demonstrates that learners tend to monitor their learning progress and adjust their learning strategies accordingly (Zimmerman, 1990; see also Thiede, Griffin, Wiley, & Redford, 2009). We generated evidence that learners do not simply rely on the current test performance and JOL but also consider them in relation to their previous experience. Reflection on the test performance and JOL at different time points altogether influences learners' selection of the subsequent study schedule.

Besides the theoretical implications, our study also offers recommendations for instruction in real educational settings. We acknowledge that it is a challenging mission to make students fully aware of and spontaneously use the ES study strategies on their own. Especially in the case of the interleaving effect, the benefit of interleaving over blocking contradicts students' intuition and prior belief, and requires extra effort to implement (e.g., see Yan et al., 2016). The present study was the first to disclose the thinking process of learners when making a spontaneous decision on a subsequent study schedule. It appears that the more confident learners feel about their own learning, the more likely they are to adopt the same way of studying. Hence, we suggest that teachers should provide their students with several opportunities to successfully study with ES learning methods. Success or failure of the earlier learning experience with a study strategy consequently affects the maintenance of that strategy (e.g., Bjorklund & Buchanan, 1989; Rabinowitz et al., 1992; Rabinowitz & McAuley, 2014; Woolley et al., 2019). Likewise, having successful initial learning experiences with an interleaved schedule may motivate students to persist in the interleaved study, resulting in better learning.

Furthermore, the present study revealed that one reason participants prefer to block their study is that they hold a misapprehension about how to master an inductive task. When the categories are highly similar to each other, it is important to find the critical differences that can help discriminate items in different categories from each other (Carvalho & Goldstone, 2014; see also Kang & Pashler, 2012). However, according to participants' written responses in the follow-up questionnaire, many of them believed that finding commonalities among the exemplars of one category was the key to succeed in the present task, which led them to pursue the less effective blocked study schedule. The present study provides an explanation of participants' nonoptimal study schedules and suggests one possible way of engaging students in the more effective interleaved schedule. That is, instructors should explain to students that recognizing the differences among categories is the key to succeeding in inductive category learning. In that way, more students would spontaneously choose to interleave their study—the schedule that facilitates the discrimination of highly similar categories.

Limitations and Future Directions

We acknowledge some limitations of the present study and suggest further research to expand our findings. First, the findings of the present study were based on a descriptive analysis of the collected data and were not directly obtained through the manipulation of the interim activity (test vs. restudy), and thus could not lead to causal inference. Further studies should manipulate the improvement on tests and/or JOLs to test the mechanism whereby an improvement of test performance and/or JOL leads to the same or a different choice of students' strategy selections. For example, the manipulation of test difficulty can affect the degree of improvement and JOL.

Second, all the participants in this research started their initial learning in an interleaved schedule in the first section, which provided them with a relatively better learning experience. Such a procedure was administered because Yan et al. (2016) reported it to be a way of helping more learners acknowledge the benefits of interleaving. For that reason, the pattern we found in the present study might or might not appear again if learners started their study with a blocked order. Yan et al. (2016) showed that starting with different study schedules resulted in different patterns of later metacognitive awareness, implying that learners had different thinking processes throughout the experiment depending on which study schedule they experienced first. Further studies would be needed to shed light on how the decision-making process changes as a function of the exposure order.

Conclusion

Though an empirically supported learning technique, interleaving is not widely endorsed and practiced by students. We have revealed that simply providing students with an interleaved study experience does not guarantee their strategy-efficacy awareness, regardless of the presence of a test-taking opportunity. What matters the most was how students viewed their own learning and how successful they felt about their current study method. Students appeared to monitor their learning progress along a continuous timeline, and successful learning experiences with interleaving in turn facilitated the adoption of the interleaved schedule for subsequent learning.

References

Anthenien, A. M., DeLozier, S. J., Neighbors, C., & Rhodes, M. G. (2018). College student normative misperceptions of peer study habit use. *Social Psychology of Education*, *21*, 303-322.

Ben-Eliyahu, A., & Bernacki, M. L. (2015). Addressing complexities in self-regulated learning: a focus on contextual factors, contingencies, and dynamic relations. *Metacognition and Learning*, *10*, 1-13.

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.

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). New York, NY: Worth Publishers.

Bjork, E. L., de Winstanley, P. A., & Storm, B. C. (2007). Learning how to learn: Can experiencing the outcome of different encoding strategies enhance subsequent encoding?. *Psychonomic Bulletin & Review*, *14*, 207-211.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417-444.

Bjorklund, D. F., & Buchanan, J. J. (1989). Developmental and knowledge base differences in the acquisition and extension of a memory strategy. *Journal of Experimental Child Psychology*, *48*, 451-471.

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging*, *3*, 249-257.

Brown, A. L., Smiley, S. S., & Lawton, S. Q. (1978). The effects of experience on the selection of suitable retrieval cues for studying texts. *Child Development*, *49*, 829-835.

Burnett, A., & Bodner, G. (2013). Generating Better Readers and/or Generators Across Two Study-Test Blocks. *Canadian Journal of Experimental Psychology*, *67*, 292-293.

Carvalho, P. F., Braithwaite, D. W., de Leeuw, J. R., Motz, B. A., & Goldstone, R. L. (2016). An in vivo study of self-regulated study sequencing in introductory psychology courses. *PloS one*, *11*(3), e0152115.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*, 481-495.

DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, *32*, 945-955.

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*, 4-58.

Flavell, J. H. (1981). Cognitive monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 35-60). New York. NY: Academic Press.

Guzman-Munoz, F. J. (2017). The advantage of mixing examples in inductive learning: A comparison of three hypotheses. *Educational Psychology*, *37*, 421-437.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement?. *Psychonomic Bulletin & Review*, *19*, 126-134.

Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97-103.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"?. *Psychological Science*, *19*, 585-592.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, *25*, 498-503.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 609-622.

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*, 203-217.

Lee, H., & 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*. 1189-1201.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, *39*, 462-476.

McCabe, J. A. (2018). What Learning Strategies Do Academic Support Centers Recommend to Undergraduates?. *Journal of Applied Research in Memory and Cognition*, *7*, 143-153.

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*, 200-219.

Morehead, K., Dunlosky, J., & Foster, N. L. (2017). Do people use category-learning judgments to regulate their learning of natural categories?. *Memory & Cognition*, *45*, 1253-1269.

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory*, *24*, 257-271.

Pastötter, B., & Bäuml, K. H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing.

Frontiers in Psychology, *5*, 286-290.

Patel, R., Liu, R., & Koedinger, K. (2016). When to block versus interleave practice? Evidence against teaching fraction addition before fraction multiplication. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2069-2074). Austin, TX: Cognitive Science Society.

Pressley, M., & Ghatala, E. S. (1989). Metacognitive benefits of taking a test for children and young adolescents. *Journal of Experimental Child Psychology*, *47*, 430-450.

Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning and Verbal Behavior*, *23*, 270-288.

Pressley, M., Levin, J. R., & Ghatala, E. S. (1988). Strategy-comparison opportunities promote long-term strategy use. *Contemporary Educational Psychology*, *13*, 157-168.

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

Rabinowitz, M., Freeman, K., & Cohen, S. (1992). Use and maintenance of strategies: The influence of accessibility to knowledge. *Journal of Educational Psychology, 84,* 211-218.

Rabinowitz, M., & McAuley, R. (2014). The effects of ease of processing on the use and perception of strategies. *Journal of Cognitive Psychology, 26,* 919-927.

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, *107*, 900-908.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481-498.

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*, 131-136.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study?. *Psychonomic Bulletin & Review*, *20*, 356-363.

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85-106). New York, NY: Routledge.

Vlach, H. A., & Kalish, C. W. (2014). Temporal dynamics of categorization: Forgetting as the basis of abstraction and generalization. *Frontiers in Psychology*, *5*, 1021-1029.

Woolley, K. E., Huang, T., & Rabinowitz, M. (2019). The effects of knowledge, strategies, and the interaction between the two in verbal analogy problem solving. *Contemporary Educational Psychology*, *56*, 91-105.

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*, 918-933.

Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, E. L., & Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied*, *23*, 403-416.

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), 8-16.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*, 3-17.

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215-221.

학습 수행 및 메타인지적 판단이 학습자의 교차학습 선택에 미치는 영향

Lan Anh Do[1], 이희승[1]

[1]연세대학교 교육학과

효과적인 학습을 위해 학습자는 학습법의 효과성을 인식하고 이를 자발적으로 실행할 수 있어야 한다. 많은 선행연구는 교차학습(모든 범주의 예를 섞어서 학습)이 묶음학습(범주별로 묶어서 학습)보다 효과적이라고 밝혔으나, 대부분의 학습자는 묶음학습이 교차학습보다 더 효과적이라고 본다. 본 연구는 그림 양식을 학습하는 과제에서 초기 학습 경험이 학습자의 후속 학습법 선택에 미치는 영향을 살펴보았다. 참가자들은 교차학습으로 첫 번째 섹션을 학습한 후 조건에 따라 중간 활동에서 시험을 보거나 재학습을 진행하였다. 그 다음, 두 번째 섹션에서 교차학습과 묶음학습 중 어떤 방식으로 학습할 것인지 스스로 선택하였다. 참가자가 선택한 방식으로 두 번째 섹션을 학습한 후 학습자들은 최종 전이 시험을 보았고, 마지막으로 후속 학습법을 선택하였다. 또한, 참가자들은 학습 후와 시험 후에 각각 자신의 학습 상태에 대한 메타인지적 판단을 내렸다. 연구 결과, 교차학습을 선택한 학습자들은 두 번째 섹션에서 더 높은 학습 수행을 보였으며, 이는 강력한 교차효과가 나타났음을 의미한다. 흥미롭게도, 초기학습에서 상대적으로 수행이 좋았던 학습자들은 후속 학습에서 더 효과적인 학습법인 교차학습을 선택하였다. 또한, 두 개 섹션에 걸쳐 시험수행 혹은 메타인지적 판단이 향상한 학습자들은 같은 학습법을 유지하고자 하였다. 이는 성공적인 학습경험이 효과적인 학습법을 선택할 수 있게 하고, 나아가 이를 지속적으로 사용하게 만드는 하나의 방법이 될 수 있음을 시사한다.

주제어: 효과적인 학습법, 교차효과, 학습법 선택, 시험