AMERICAN PSYCHOLOGICAL ASSOCIATION SARMAC Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681

2024, Vol. 13, No. 3, 393–406
https://doi.org/10.1037/mac0000194

EMPIRICAL ARTICLE

Interleaved Pretesting Enhances Category Learning and Classification Skills

Steven C. Pan[1], Ganeash Selvarajan[1], and Chanda S. Murphy[2]
[1] Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Singapore
[2] Department of Psychological Science and Counseling, Austin Peay State University, United States

OPEN DATA OPEN MATERIALS PREREGISTERED+ TC

Alternating between concepts during learning (*interleaving*) and making guesses about to-be-learned information before viewing the correct answers (*pretesting*) can enhance learning relative to focusing on one concept at a time (blocking) and studying, respectively. We investigated the potential benefits of interleaving and pretesting for acquiring categorical knowledge and classification skills. In three experiments, participants learned about psychopathological disorders from interleaved or blocked case studies and via pretesting or studying. A 5-min delayed test (Experiment 1) showed that interleaving and pretesting improved the ability to classify new and previously viewed case studies. Moreover, their combination had at least additive effects, yielding the best overall performance. Similar results occurred on a 48-hr delayed test (Experiment 2) and under conditions of equivalent time on task (Experiment 3). Overall, this study reveals that an effective scheduling approach paired with a beneficial learning activity forms a potent combination (*interleaved pretesting*) that is uniquely capable of enhancing learning.

General Audience Summary

A growing body of research suggests that alternating between multiple concepts or topics during learning, which is also known as *interleaving*, is more effective than a traditional, one-topic-at-a-time approach, which is also known as blocking. There is also evidence that making guesses about information before studying the correct answers, or *pretesting*, can yield better learning than studying correct information without any guessing. This study investigated (a) the benefits of interleaving and/or pretesting for learning to identify different categories and (b) whether combining interleaving and pretesting—that is, viewing a series of concepts in an interleaved order and guessing the identity of each concept prior to learning the correct answer—might also improve learning. In each of the three experiments, adult participants learned to identify psychopathological disorders (e.g., cyclothymic disorder) via exposure to paragraph-long case study examples of each disorder. Learning was interleaved or blocked (alternating between concepts or focusing on one concept at a time) and involved pretesting or studying (guessing the identity of the disorder first or being told the specific disorder from the outset). After 5 min (Experiments 1 and 3) or 48 hr (Experiment 2), participants took a classification test wherein they had to identify disorders described in never-before-seen and previously viewed case studies. If learning involved interleaving or pretesting, then classification performance was improved relative to blocking or studying, respectively. If interleaving and pretesting were used together, then classification performance was even better. These results suggest that combining an effective scheduling approach (interleaving) and a beneficial learning activity (pretesting) can harness benefits of both strategies, yielding better learning than when either strategy is used alone. Accordingly, in analogous situations and possibly other contexts, learners stand to benefit most from using interleaving and pretesting in tandem.

Keywords: interleaving, interleaved practice, pretesting, prequestioning, category learning

Supplemental materials: https://doi.org/10.1037/mac0000194.supp

Steven C. Pan https://orcid.org/0000-0001-9080-5651
Ganeash Selvarajan https://orcid.org/0000-0002-4768-1793
Chanda S. Murphy https://orcid.org/0009-0000-6770-078X

This study's design and hypotheses were preregistered at https://AsPredicted.org, accessible at https://aspredicted.org/QV1_FCY for Experiment 1, https://aspredicted.org/C7R_PDK for Experiment 2, and https://aspredicted.org/WR9_W1S for Experiment 3. Data and example stimuli are

continued

Of the numerous learning strategies that have been investigated to date, some of the most promising can be divided into two classes: (a) *scheduling study*, which details how learning should be arranged optimally in time (i.e., "when" or "in what order" to learn), and (b) *learning activity*, which specifies how a learner might interact with the material (i.e., "what one should do" while learning; for reviews, see Carpenter et al., 2022; Dunlosky et al., 2013). An example of (a) is *interleaving*, which involves alternating between a series of to-be-learned concepts or topics (Rohrer, 2012), whereas an example of (b) is *pretesting*, which involves making guesses about to-be-learned information before viewing the correct answers (Richland et al., 2009). Educationally relevant uses of both strategies, including in isolation and in combination, are the focus of this article. Both interleaving and pretesting are potential "desirable difficulties"—that is, evidence-based learning strategies that are more challenging and error prone when first used, but ultimately better for learning (Bjork & Bjork, 2011).

Historically, learning strategy research has focused on strategies in isolation, comparing one strategy versus a "business-as-usual" or control condition. More recently, however, interest has grown in examining the combined effects of multiple effective strategies (e.g., Rawson & Dunlosky, 2011; see also Y. Kang et al., 2023; McDaniel, 2023; Pan et al., 2024; Roelle et al., 2023). For instance, combining spacing and retrieval practice can enhance concept learning and other outcomes (Rawson & Dunlosky, 2022). Conversely, combining strategies is sometimes counterproductive (e.g., spacing and interleaving in Birnbaum et al., 2013). Identifying the learning processes and potential benefits of combined strategies can facilitate their effective use with appropriate learning materials and situations (Roelle et al., 2023). Combining strategies may yield different outcomes: redundancy or counterproductivity may yield no added benefits, while additive, interactive, or synergistic advantages may enhance learning.

The Interleaving Effect and the Pretesting Effect

Interleaving, wherein learners alternate between different concepts or topics, can improve learning compared to the traditional approach of focusing on one concept at a time (blocking). This improvement, known as the *interleaving effect*, is most often observed in inductive category learning (i.e., learning from examples) and especially when the to-be-learned categories are highly similar and confusable (for reviews, see Brunmair & Richter, 2019; Carpenter & Pan, 2024; Carvalho & Goldstone, 2019; S. H. K. Kang, 2017; Rohrer, 2012). For example, Zulkiply et al. (2012, cf. Murphy & Pavlik, 2018) had undergraduate students learn psychopathological disorders through interleaved case studies, wherein each successive case study represented a different disorder, or blocked case studies, wherein studies referencing the same disorder were grouped together. On a subsequent classification test requiring identification of disorders from new case studies, performance was better following interleaving than blocking.

Theoretical explanations for the interleaving effect differ, with the two primary accounts focusing on temporal spacing and discriminative contrast. The former suggests that the interleaving effect is a manifestation of the well-established spacing effect (Ebbinghaus, 1885), wherein increased time between exposures to stimulus materials enhances learning (Carpenter et al., 2022; Cepeda et al., 2006; Jacoby et al., 2010). The latter proposes that interleaving causes learners to compare differences between categories, thereby enhancing learning (Birnbaum et al., 2013; S. H. K. Kang & Pashler, 2012; see also Carvalho & Goldstone, 2019). Although evidence both supports and challenges these accounts, research involving visual stimuli generally favors the discriminative contrast explanation (e.g., Birnbaum et al., 2013), while studies with nonvisual materials have sometimes favored a spacing-based account (e.g., Foster et al., 2019).

Pretesting, also known as errorful generation or, in certain contexts, prequestioning, entails making guesses about to-be-learned information before studying the correct answers. In a variety of contexts (e.g., with text or video materials), it can enhance learning compared to studying correct information, a phenomenon known as the *pretesting effect* (for a review, see Pan & Carpenter, 2023). For example, Richland et al. (2009) found that having participants engage in pretesting before reading a text passage about achromatopsia, which yielded many incorrect guesses, resulted in better comprehension test performance than reading without pretesting.

The pretesting effect has been the focus of various theoretical accounts. In paired associate learning, it has been suggested that pretesting fosters mediator generation (Huelser & Metcalfe, 2012), activates a search set of possible answers (Grimaldi & Karpicke, 2012), forms episodic memories of generating errors and learning the correct answers (Jacoby & Wahlheim, 2013), and/or activates an error signal

publicly available in the Open Science Framework and can be accessed at https://osf.io/jf298/. Portions of this research were presented at Ruhr-Universität Bochum and the Würzburg International Autumn School on Lasting Learning in October 2023, plus the inaugural National University of Singapore Undergraduate Research Opportunities Programme Research Experience Conference in January 2024.

The authors have no conflicts of interest to disclose. This research was supported by a National University of Singapore Research Experience Programme Grant (UROP-REx) from the Undergraduate Research Opportunities Programme awarded to Ganeash Selvarajan under the supervision of Steven C. Pan. The authors thank Benedict Yew for the assistance with data processing and Michelle Kaku for the assistance with online data collection. The authors also thank Roman Abel for the suggestions regarding data analysis.

Steven C. Pan played a lead role in conceptualization, formal analysis, methodology, resources, supervision, visualization, writing–original draft, and writing–review and editing, a supporting role in funding acquisition, and an equal role in investigation. Ganeash Selvarajan played a lead role in data curation, funding acquisition, and project administration, a supporting role in conceptualization, formal analysis, writing–original draft, and writing–review and editing, and an equal role in investigation and methodology. Chanda S. Murphy played a supporting role in conceptualization, investigation, writing–original draft, and writing–review and editing.

The data are available at https://osf.io/jf298.

The experimental materials are available at https://osf.io/jf298.

The preregistered design and analysis plan (transparent changes notation) is accessible at https://aspredicted.org/QV1_FCY (Experiment 1), https://aspredicted.org/C7R_PDK (Experiment 2), and https://aspredicted.org/WR9_W1S (Experiment 3).

Correspondence concerning this article should be addressed to Steven C. Pan, Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, 9 Arts Link, Singapore 117572, Singapore. Email: scp@nus.edu.sg

(S. H. K. Kang et al., 2011), any of which may cause the pretesting effect. Additionally, pretesting may boost interest, curiosity, and attention, as well as prompt a search for correct answers (Pan et al., 2020; Rothkopf, 1966; Sana & Carpenter, 2023). To date, there is a range of evidence supporting each of these explanations (Mera et al., 2022; Pan & Carpenter, 2023).

Are Interleaving and Pretesting Complementary Learning Strategies?

What is the effectiveness of a strategy for scheduling study, interleaving, and a learning activity strategy, pretesting, for acquiring categorical knowledge and classification skills pertaining to psychopathological disorders? As previously noted, presenting case studies in an interleaved fashion can improve classification skills compared to blocked learning (Zulkiply et al., 2012). Whether pretesting enhances learning in this context (e.g., if learners had to guess the type of disorder before learning the correct answer), however, has not previously been investigated. Moreover, the potential benefits of combining interleaving and pretesting (e.g., by presenting case studies in an interleaved order *and* engaging in pretesting beforehand) have yet to be established.

Using interleaving and pretesting, in our estimation, could yield several possible outcomes. First, although prior research suggests that interleaving can be beneficial, it has been unclear whether pretesting will be helpful as well. In visual category learning research, however, guessing the category membership of example shapes before receiving correct answer feedback can enhance classification skills relative to studying examples (e.g., Carvalho & Goldstone, 2015; cf. Choi & Lee, 2020), particularly when the categories cannot be learned using simple rules (Ashby et al., 2002). For psychopathological disorders presented in text format, pretesting might aid learning in analogous fashion by engaging diagnostic processes necessary for classification tests, stimulating engagement with correct answers, or other means (Pan et al., 2020; Sana & Carpenter, 2023). If so, then pretesting might also be beneficial.

As for the combination of interleaving and pretesting, we considered at least three possibilities (cf. Rawson & Dunlosky, 2011). First, one strategy might negatively affect the other. For example, if the interleaving effect relies on uninterrupted discriminative contrast (as suggested by Birnbaum et al., 2013, and others), then interrupting that process (as might occur if the process of generating guesses disrupts the comparison of case studies) could be detrimental. Second, interleaving and pretesting might be partially or fully redundant, particularly if they engage the same or overlapping cognitive processes. For example, both strategies might cause learners to focus on distinguishing features that differentiate categories (Carvalho & Goldstone, 2015). Alternatively, although both strategies may focus attention on distinguishing features, pretesting may also cause the generation or enhancement of relevant memories; hence, using the two strategies could yield subadditive benefits over using either strategy alone. Finally, additive or even synergistic benefits may occur, particularly if wholly different mechanisms are involved or different aspects of the learning process are enhanced.

The Present Study

This study entailed three experiments. In each experiment, participants learned psychopathological disorders in a *blocked studying*, *interleaved studying*, *blocked pretesting*, or *interleaved pretesting* group. These groups formed a factorial design, allowing us to address the independent effects of interleaving and pretesting, as well as their combination. To gain insights into participants' reactions to the strategies used and the resulting learning experiences, we also solicited metacognitive judgments. Then, after a 5-min (Experiments 1 and 3) or 48-hr (Experiment 2) retention interval, participants completed a classification test that involved identifying disorders from new and previously presented case studies.

Experiment 1

In the first experiment, participants learned about six psychopathological disorders from three example case studies of each disorder. Learning was interleaved (i.e., wherein examples of multiple disorders were mixed) or blocked (i.e., wherein all examples of each disorder were grouped together) and prefaced by pretesting (i.e., attempting to guess the disorder that was presented) or no pretesting at all (i.e., studying only).

Method

The entire study was conducted online using Qualtrics. The design, hypotheses, sampling strategy, and analysis plan were preregistered at https://aspredicted.org/QV1_FCY. Determination of sample size, all data exclusions, all manipulations, and all measures are detailed below.

Participants

The target sample size was determined via a priori power analysis conducted in G*Power (Faul et al., 2007). That power analysis indicated that at least 32 participants per group are needed for 80% power to detect a medium-small effect size of Cohen's $f = 0.25$ in a 2×2 between-participants design at $\alpha = .05$. We recruited more than that number per group. One-hundred seventy participants were recruited online via Prolific Academic in exchange for a payment of GBP £4.20 or USD $5.26 per participant. All participants had to be from an English-speaking country (i.e., Australia, Canada, New Zealand, the United Kingdom, or the United States), be fluent in English, be aged between 21 and 40 years, have an approval rate of 95% or higher on prior Prolific studies, not been personally diagnosed or have any close friends or family members that have been diagnosed with psychopathological disorders, and not formally learned about such disorders. Geographical, approval rate, and age requirements were specified within the study's Prolific listing, whereas prior experience and knowledge of disorders were established via a series of screening questions posed at the outset of the experiment.

Prior to formal analysis, data from one and 20 participants, respectively, were excluded for incomplete responding and evidence of off-task browser activity (as detected via TaskMaster, which was programmed into the experiment; Permut et al., 2019). In alignment with our preregistered exclusion criteria, data from 23 participants were also excluded for classification test scores that were two or more standard deviations above or below mean group performance. The participants excluded from data analysis were not dominated by any specific group. The final sample, which totaled 128 participants (*blocked studying* group, $n = 32$; *interleaved studying* group, $n = 33$; *blocked pretesting* group, $n = 32$; *interleaved pretesting* group, $n = 29$), had a mean age of 32.7 years and was 56% male. Sixty percent of these participants were from the United Kingdom, 15% were from the United States, and 25% were from the remaining eligible countries; 13% of participants were Asian, 10% were Black, 8% were mixed, 65% were White, and 4% were from other ethnic groups or declined to provide ethnicity information.

The entire study was conducted with ethics board approval obtained at the first and second authors' affiliated university. All participants provided informed consent prior to experimentation and were treated in accordance with the principles set out in the Declaration of Helsinki.

Design

This experiment featured a 2 × 2 between-participants factorial design with factors of Training Schedule (Blocked vs. Interleaved) and Training Activity (Studying vs. Pretesting).

Materials

The materials consisted of 30 case studies based on Zulkiply et al. (2012) and similar to those used by Murphy (2017) and Murphy and Pavlik (2018). There were five case studies for each of six psychopathological disorders (attention-deficit hyperactivity disorder, autism spectrum disorder, bipolar disorder, borderline personality disorder, intellectual development disorder, and schizophrenia). Each case study consisted of a single paragraph of 100–120 words in length describing an individual with behavioral characteristics that met the diagnostic criteria for the respective disorder according to the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition, from the American Psychiatric Association (see Appendix for examples). Three case studies per disorder were used during training. Two of these case studies were repeated on the subsequent classification test, whereas an additional two case studies per disorder appeared only on the classification test.

To minimize the common name of a disorder providing a hint for characteristics of the disorder itself, each disorder's name was replaced with a less common name (i.e., cyclothymic affect disorder, dysfunctional cognition disorder, pervasive development disorder, resonance development disorder, schismic cognition disorder, and self-regulation disorder). This approach, which retained a degree of clinical accuracy with the materials while making those materials more challenging to learn, constituted an intermediate alternative to Zulkiply et al.'s (2012) use of nonsense names and Murphy and Pavlik's (2018) use of common names.

Procedure

At the outset of the experiment, each participant was randomly assigned by computer to the blocked studying, interleaved studying, blocked pretesting, or interleaved pretesting group. All participants gave informed consent and completed the initial screening questions. They then underwent a training phase, answered metacognitive questions, took a short break, and finally completed a classification test.

Training Phase. Group assignment determined how the disorders were learned. A depiction of the approach to ordering case studies in the blocked and interleaved groups is presented in Table 1.

Blocked Studying and Interleaved Studying. Participants were informed that they would view case studies of psychopathological disorders and that their goal was to learn the characteristics of each disorder. In the interleaved studying group, participants were further told that the case studies would be shown in random order. As such, they should expect to encounter case studies for a given disorder at different points during the training phase (this clarification was included to ensure that participants did not mistake the interleaved sequence for a malfunctioning study).

After participants had finished reading the instructions, the case studies were presented one at a time for 40 s each (as in Murphy & Pavlik, 2018, and 10 s longer than in Zulkiply et al., 2012). With each case study, the name of the disorder was shown in bold font above the case study text. In the blocked studying condition, all three case studies per disorder were presented in succession without any intervening case studies addressing other disorders. The presentation of case studies for each disorder was organized into blocks, wherein each block consisted of three different examples of the same psychopathological disorder. Thus, each disorder was learned in a single block and not revisited at any other point.

In the interleaved studying group, the case studies were organized into three blocks of six case studies each, with only one case study per disorder within each block. Each block presented the case studies in an interleaved order with the constraint that the last case study within each block could not involve the same disorder as the first case study for the next block. Further, to meet that constraint, the interleaved pattern within each block relied on a fixed, predetermined pattern rather than a fully randomized pattern (cf. Pan et al., 2019). Across all 18 presented case studies, each successive case study involved a different disorder than the one that had just been shown, and moreover, no disorder was shown more than once within each block of six case studies. Overall, within each of the three blocks, participants saw one case study from each disorder in an interleaved pattern.

Blocked Pretesting and Interleaved Pretesting. Participants were informed that they would be shown case studies of psychopathological disorders, one case study at a time. The disorders would initially not be identified by name; rather, participants would have to read the case study and, within 40 s, guess the disorder that it best represented. The names of the six to-be-learned disorders were provided for participants to choose from and register their guess. After the allotted time had elapsed, the correct answer appeared in

Table 1
Example Case Study Training Schedules

Group	Example arrangement
Blocked studying, blocked pretesting	A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3, D_1, D_2, D_3, E_1, E_2, E_3, F_1, F_2, F_3
Interleaved studying, interleaved pretesting	A_1, B_1, C_1, D_1, E_1, F_1, A_2, B_2, C_2, D_2, E_2, F_2, A_3, B_3, C_3, D_3, E_3, F_3

Note. Letters represent psychopathological disorders, and subscripts represent case study numbers. Example arrangement has been simplified for ease of exposition.

bold text next to the case study. Participants were told to read the correct answer and think about why the case study best represented that disorder. To verify that they had viewed the correct answer, they were also required to indicate whether they had identified the case study or not (via a yes/no question). After doing so, they were permitted to advance to the next case study.

To alleviate potential concerns about the difficulty of identifying disorders, participants were told that guessing incorrectly was acceptable and that making guesses would likely be challenging at first. As more case studies were presented and the disorders became more familiar, however, it was likely that they would be able to make more accurate guesses.

The ordering of case studies in the blocked pretesting and interleaved pretesting groups resembled the ordering of case studies in the blocked studying and interleaved studying groups, respectively. In addition, just as in the interleaved studying group, participants in the interleaved pretesting group were told that the case studies would be presented in random order.

Metacognitive Questions and Short Break. In each group, after all 18 case studies had been presented, participants made a global judgment of learning ("How confident are you that you have learnt all the concepts of mental disorders presented in this study?") and a global judgment of difficulty ("How difficult was it for you to learn the concepts of mental disorders presented in this study?"), both using a 1–10 Likert sliding scale. Next, they made two predictions of future test performance (in terms of predicted percentage correct). The first prediction involved a hypothetical test wherein they had to identify never-before-seen case studies, each exemplifying one of the disorders that they had just learned, whereas the second prediction involved another hypothetical test wherein they had to identify the disorders represented by the exact same case studies that they had just seen. The purpose of the metacognitive questions was to explore participants' perceptions of the learning experience conferred by the different approaches used.

After answering the metacognitive questions, participants were instructed to rest their eyes. A countdown timer of 30 s was shown, and participants were allowed to proceed after at least 30 s' rest. Overall, the total amount of time spent answering the metacognitive questions, plus the short break, yielded a retention interval of approximately 5 min.

Classification Test. At the end of the study, all participants completed a classification test featuring two sections of 12 case studies each. The first section involved never-before-seen case studies, two per disorder, and the second section involved previously viewed case studies, two per disorder. Prior to each section, participants were instructed to apply what they had learned to classify the case studies being presented. In each section, case studies were presented one at a time in random order. The name of the disorder that the case study represented was not shown; rather, for each case study, participants were required to identify the presented disorder from a list of the names of the six previously learned disorders. They were given unlimited time to respond, and once they had entered the answer, the next case study was shown. No feedback was provided.

To prevent any effects of reexposure to previously viewed case studies from affecting classification performance for new case studies, the new case studies were always presented before previously viewed case studies. After participants had finished responding to all 24 case studies on the classification test, they were debriefed, dismissed, and compensated for their participation.

Results

Pretest Performance

Mean guessing performance for the blocked pretesting and interleaved pretesting groups is presented in the upper panel of Figure 1. As shown in the figure, performance in the blocked pretesting group generally rose across each block of three case studies, then dropped at the start of the next block (i.e., case studies 4, 7, 10, 13, and 16). The interleaved pretesting group also varied in guessing performance but did not exhibit oscillations that were as pronounced or systematic, particularly in the latter half of the training phase. Overall, guessing performance improved with practice in both groups, although the magnitude of that improvement was larger with blocking versus interleaving. In the blocked pretesting group, accuracy improved from the first case study ($M = 0.46$, $SE = 0.036$) to the third case study ($M = 0.79$, $SE = 0.095$) representing each disorder (~0.33 improvement), $t(31) = 5.78$, $p < .0001$, $d = 1.022$, and in the interleaved pretesting group, accuracy improved from the first case study ($M = 0.61$, $SE = 0.040$) to the third case study ($M = 0.78$, $SE = 0.11$) representing each disorder (~0.17 improvement), $t(28) = 3.51$, $p = .0015$, $d = 0.65$.

A noticeable pattern is that guessing performance for the very first of the 18 case studies was higher in the interleaved pretesting group than the blocked pretesting group. That result likely stems from the implementation of fixed interleaved patterns within each six-trial block to meet design constraints (post hoc analysis revealed that an unintended consequence of that design decision was that the very first case study that was presented in the interleaved pretesting group tended to be among the easiest to guess; that ease of guessing was however not maintained and subsequent case studies were more difficult to correctly guess). Importantly, a comparison of specific case studies did not show any substantial between-group differences in overall guessing accuracy for any disorder.

Time on Task

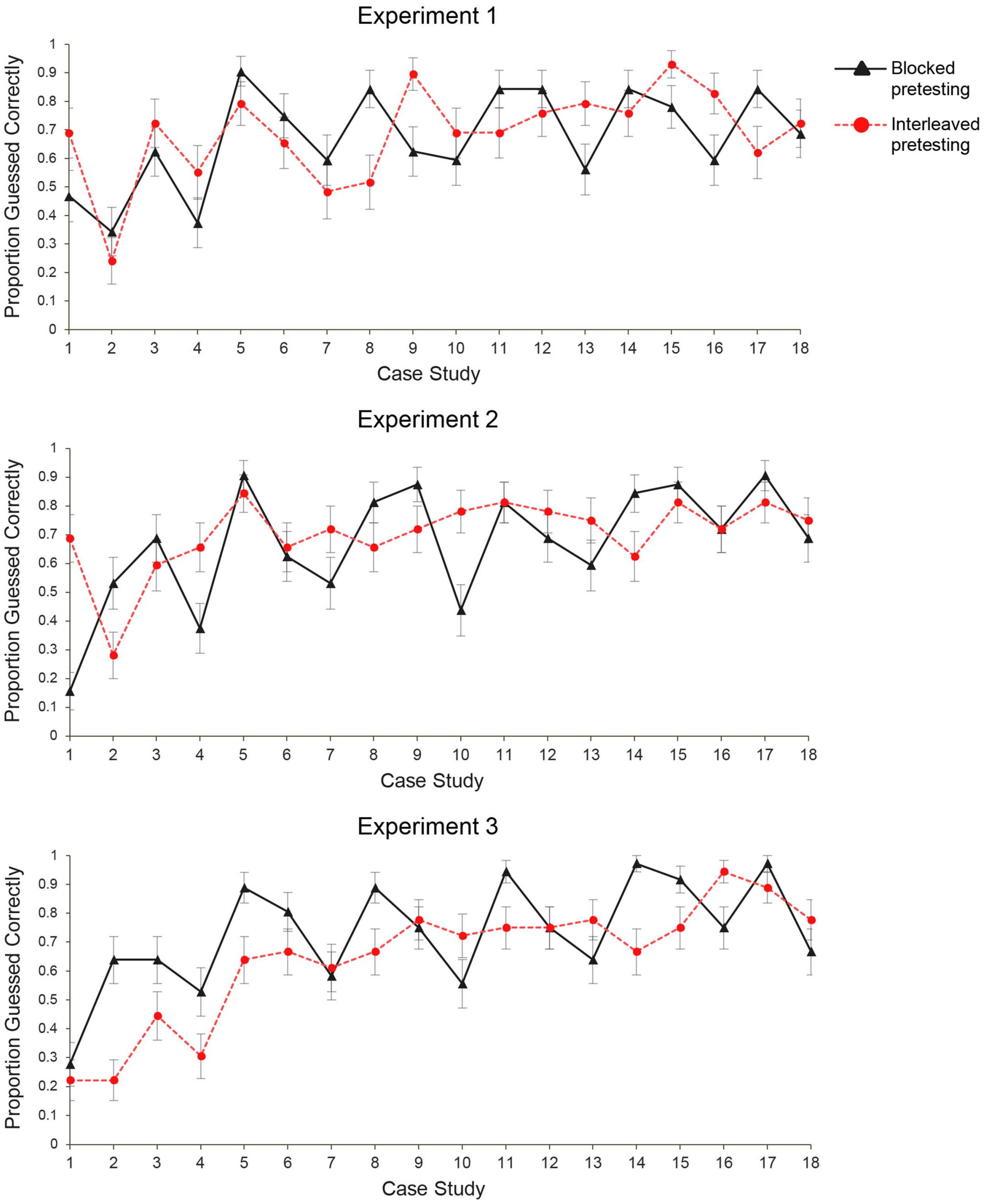
Whereas the time per case study was exactly 40 s in the blocked studying and interleaved studying groups, participants were allotted extra time to read the correct answer once it was presented and answer a verification question regarding whether their guess had been correct or not. Doing so resulted in additional time per case study of $M = 21.4$ s, $SE = 1.8$ s, in the blocked pretesting group and $M = 19.3$ s, $SE = 1.1$ s, in the interleaved pretesting group.

Classification Test Performance

Classification test results involving new and previously viewed case studies are displayed in the upper panel of Figure 2. In accordance with our preregistered analysis plan, classification test results for new case studies and previously viewed case studies were analyzed separately. In a departure from that analysis plan, however, factorial analyses of variance (ANOVAs), which are more appropriate for a

Figure 1

Training Phase Performance for Each Pretesting Group

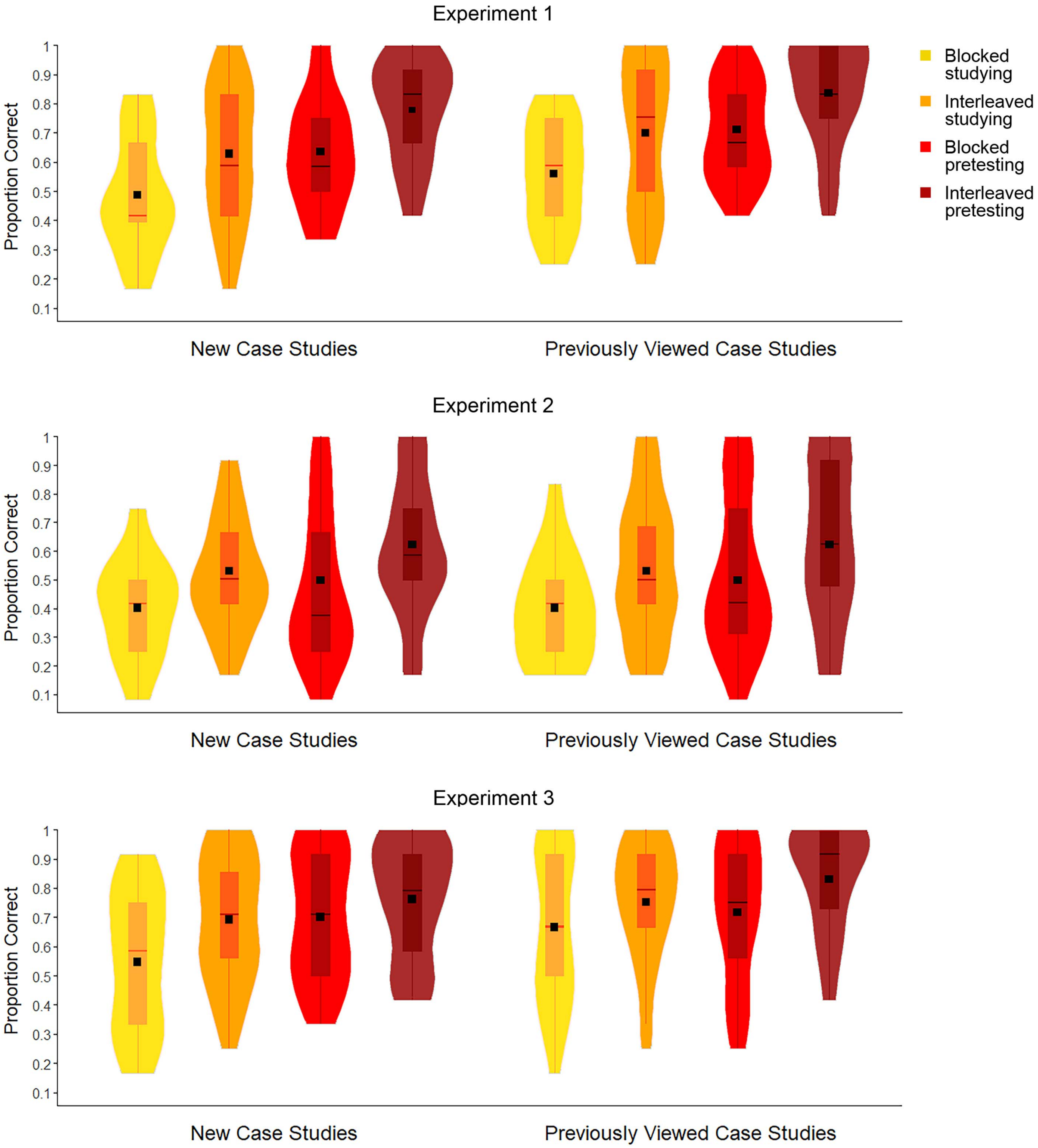


Note. Results shown in chronological order (i.e., starting with the first case study encountered by each participant and ending with the 18th and last case study that was encountered). Error bars = standard error of the mean. See the online article for the color version of this figure.

study featuring a 2 × 2 factorial design, were performed rather than a one-way ANOVA (that oversight was corrected in the preregistration for Experiment 2).

New Case Studies. A 2 × 2 ANOVA with factors of Training Schedule (Blocked vs. Interleaved) and Training Activity (Studying vs. Pretesting) conducted on participant-level mean test scores for new case studies revealed a significant effect of Training Schedule, $F(1,\ 122) = 14.86$, $p < .001$, $\eta_p^2 = 0.11$; a significant effect of Training Activity, $F(1,\ 122) = 17.40$, $p < .0001$, $\eta_p^2 = 0.12$; and no significant interaction ($p = .98$). Those results align with inspection of Figure 2 (upper panel, left side), wherein there are indications of an advantage of interleaving over blocking, pretesting versus

Figure 2
Classification Test Results



Note. The classification test occurred after a delay of 5 min (Experiments 1 and 3) or at least 48 hr (Experiment 2). Error bars = standard error of the mean. See the online article for the color version of this figure.

studying, and best overall performance in the interleaved pretesting group; moreover, it appears that the interleaved studying and blocked pretesting groups performed similarly.

Previously Viewed Case Studies. A 2 × 2 ANOVA analogous to that performed for new case study data revealed a significant effect of Training Schedule, $F(1, 122) = 13.35$, $p < .001$, $\eta_p^2 = 0.099$; a significant effect of Training Activity, $F(1, 122) = 16.86$, $p < .0001$, $\eta_p^2 = 0.12$; and no significant interaction ($p = .84$). Those results are consistent with visual inspection of Figure 2 (upper panel, right side), wherein there are indications of an advantage of interleaving over blocking, pretesting versus studying, and best overall performance in the interleaved pretesting group.

Metacognitive Questions

Results for the metacognitive questions are presented in Table 2. We performed separate 2 × 2 ANOVAs for each question type with factors of Training Schedule (Blocked vs. Interleaved) and Training Activity (Studying vs. Pretesting). For judgments of learning, judgments of difficulty, and predictions of future test performance for new cases, there were no significant main effects or interactions involving Training Schedule or Training Activity (all $ps \geq .076$). Numerically, however, there were signs that blocked studying was rated as the easiest and other training methods, especially interleaved pretesting, were rated as more difficult. Further, a 2 × 2 ANOVA performed on predictions of future test performance for previously viewed cases revealed a significant main effect of Training Activity, $F(1, 122) = 7.31$, $p = .0078$, $\eta_p^2 = 0.057$, and no significant main effect of Training Schedule or interaction ($ps \geq .44$). That main effect reflects higher predictions from participants in the pretesting versus studying groups.

Experiment 2

In the first experiment, an interleaving effect for learning about psychopathological disorders was observed, which replicates prior research (cf. Pan et al., 2024; Zulkiply & Burt, 2013). There was also evidence of a pretesting effect, which establishes that pretesting can be beneficial for learning of such materials. Moreover, interleaved pretesting yielded the highest scores, suggesting that combining interleaving and pretesting is more effective than either strategy alone. Emerging evidence indicates that pretesting effect magnitude can increase after a longer retention interval (Kliegl et al., 2024). Accordingly, a second experiment was conducted to conceptually replicate the previous results and examine learning patterns after a delay of at least 2 days. Nearly all other aspects of the design and procedure remained the same.

Method

This experiment was preregistered at https://aspredicted.org/C7R_PDK.

Participants

The target sample size and sampling methods followed that of the first experiment. One-hundred fifty-one participants were recruited via Prolific Academic in exchange for a payment of at least GBP £2.10 or USD $2.63 per participant (an additional £2.10 or $2.63 bonus was offered to participants that completed both parts of the experiment). Prior to formal analysis, data were removed from 11 participants for evidence of off-task browser activity, two participants that did not follow study instructions, three participants for classification test scores that were two or more standard deviations away from mean group performance, and nine participants that did not complete the classification test within 72 hr after being asked to do so. The final sample consisted of 126 participants (*blocked studying* group, $n = 30$; *interleaved studying* group, $n = 32$; *blocked pretesting* group, $n = 32$; *interleaved pretesting* group, $n = 32$), had a mean age of 32.9 years, and was 53% male. Sixty-five percent of these participants were from the United Kingdom, 16% were from Canada, 10% were from Australia, and 9% were from the remaining eligible countries; 20% of participants were Asian, 10% were Black, 2% were mixed, 67% were White, and 2% were from other ethnic groups or declined to provide ethnicity information.

Design, Materials, and Procedure

All aspects of the design, materials, and procedure were identical to the preceding experiment except for the following changes. First, for logistical reasons, the screening for prior knowledge and experience was conducted prior to, rather than at the start of, the experiment. Doing so entailed presenting the screening questions via a separate study listing on Prolific that participants had to complete beforehand. Second, the classification test was delayed by at least 2 days. When 48 hr had elapsed from the release of the first part on Prolific Academic, participants were sent the link to the classification test and allotted up to 72 hr to complete it. Due to the time difference between Singapore (where the research team is headquartered) and countries where many participants were located, however, the time at which the second part was sent and completed by participants varied. The absolute upper and lower limits of the retention interval between the first and second parts were 40 and 72 hr, respectively. (On average, participants

Table 2
Metacognitive Data

		Judgment, *M* (*SE*), from 1 to 10		Predicted test result, *M* (*SE*), in %	
Experiment	Group	Learning	Difficulty	New case study	Previously viewed case study
1	Blocked studying	5.1 (0.4)	5.0 (0.3)	52.5 (3.5)	65.2 (3.3)
	Interleaved studying	4.6 (0.3)	5.6 (0.4)	51.1 (3.0)	62.6 (2.3)
	Blocked pretesting	5.4 (0.4)	5.8 (0.4)	55.4 (3.1)	73.5 (2.4)
	Interleaved pretesting	5.1 (0.5)	6.2 (0.4)	57.7 (4.8)	72.4 (5.2)
2	Blocked studying	4.9 (0.3)	5.2 (0.4)	48.6 (3.8)	60.1 (1.7)
	Interleaved studying	5.5 (0.3)	6.4 (0.3)	57.8 (3.5)	67.7 (1.8)
	Blocked pretesting	5.7 (0.3)	6.3 (0.4)	61.4 (3.5)	77.5 (2.0)
	Interleaved pretesting	6.2 (0.4)	5.9 (0.4)	62.2 (3.4)	78.6 (3.5)
3	Blocked studying	5.2 (0.3)	5.4 (0.3)	54.5 (3.2)	66.8 (2.9)
	Interleaved studying	5.7 (0.3)	5.6 (0.3)	54.8 (3.8)	66.5 (3.2)
	Blocked pretesting	6.0 (0.3)	6.9 (0.2)	60.8 (3.1)	79.5 (2.4)
	Interleaved pretesting	6.8 (0.3)	6.1 (0.3)	68.7 (2.7)	81.8 (2.7)

Note. *SE* = standard error.

completed the second part of the experiment 51 hr after completing the first part; the retention interval did not significantly differ between groups.)

Results

Pretest Performance

Mean guessing performance for the blocked pretesting and interleaved pretesting groups is presented in the middle panel of Figure 1. The same overall patterns as in the prior experiment were observed, including more pronounced and systematic oscillations in the blocked pretesting group than in the interleaved pretesting group. Guessing performance also improved with practice in both groups, although the magnitude of that improvement was again larger in the case of blocking versus interleaving. In the blocked pretesting group, guessing accuracy improved from the first case study (M = 0.39, SE = 0.024) to the third case study (M = 0.81, SE = 0.10) of each disorder (~0.42 improvement), $t(31) = 9.55$, $p < .0001$, $d = 1.69$, and in the interleaved pretesting group, guessing accuracy also improved from the first case study (M = 0.62, SE = 0.049) to the third case study (M = 0.74, SE = 0.15) of each disorder (~0.12 improvement), $t(31) = 2.041$, $p = .050$, $d = 0.36$.

Time on Task

As in the prior experiment, participants in the pretesting groups spent additional time per case study than the blocked studying and interleaved studying groups, who spent exactly 40 s studying each case study (additional time of M = 19.7 s, SE = 1.9 s, in the blocked pretesting group and M = 19.8 s, SE = 1.6 s, in the interleaved pretesting group).

Classification Test Performance

Classification test results involving new and previously viewed case studies are displayed in the middle panel of Figure 2. In accordance with the preregistered analysis plan, classification test results for new case studies and previously viewed case studies were analyzed separately using factorial ANOVAs.

New Case Studies. A 2 × 2 ANOVA with factors of Training Schedule (Blocked vs. Interleaved) and Training Activity (Studying vs. Pretesting) conducted on participant-level mean test scores for new case studies revealed a significant effect of Training Schedule, $F(1, 122) = 9.79$, $p = .0022$, $\eta_p^2 = 0.074$; a significant effect of Training Activity, $F(1, 122) = 4.14$, $p = .044$, $\eta_p^2 = 0.033$; and no significant interaction ($p = .83$). Those results are consistent with inspection of Figure 2 (middle panel, left side), wherein there are indications of an advantage of interleaving over blocking, an advantage for pretesting versus studying, and best overall performance in the interleaved pretesting group.

Previously Viewed Case Studies. A 2 × 2 ANOVA with factors of Training Schedule (Blocked vs. Interleaved) and Training Activity (Studying vs. Pretesting) conducted on participant-level mean test scores for previously viewed case studies revealed a significant effect of Training Schedule, $F(1, 122) = 8.41$, $p = .0044$, $\eta_p^2 = 0.064$; a significant effect of Training Activity, $F(1, 122) = 6.08$, $p = .015$, $\eta_p^2 = 0.047$; and no significant interaction ($p = .74$). Those results are consistent with inspection of Figure 2 (middle panel, right side), wherein there were indications of an advantage of interleaving over blocking, pretesting versus studying, and best overall performance in the interleaved pretesting group.

Metacognitive Questions

Data from the metacognitive questions (Table 2) were analyzed using 2 × 2 ANOVAs as in the preceding experiment. The following significant results were obtained; all other main effects or interactions not mentioned here were not significant. For judgments of learning, there was a significant main effect of Training Activity, $F(1, 122) = 4.41$, $p = .038$, $\eta_p^2 = 0.035$, reflecting higher judgments in the pretesting versus studying groups. For judgments of difficulty, there was a significant Training Schedule × Training Activity interaction, $F(1, 122) = 5.16$, $p = .025$, $\eta_p^2 = 0.041$, reflecting higher difficulty ratings for interleaving versus blocking in the studying groups but not in the pretesting groups. For predictions of future test performance for new cases, there was a significant main effect of Training Activity, $F(1, 122) = 5.85$, $p = .017$, $\eta_p^2 = 0.46$, as was observed for predictions of future test performance for previously viewed cases, $F(1, 122) = 18.93$, $p < .0001$, $\eta_p^2 = 0.13$. Those differences reflected higher predictions from participants in the pretesting versus studying groups.

Experiment 3

The third experiment investigated the reproducibility of the observed patterns under conditions of equal time on task.

Method

This experiment was preregistered at https://aspredicted.org/WR9_W1S.

Participants

The target sample size and sampling methods followed that of the preceding experiments. One-hundred eighty-two participants were recruited from Prolific Academic in exchange for a payment of at least GBP £4.20 or USD $5.26 per participant. Prior to formal analysis, data were removed from 17 participants for evidence of substantial off-task browser activity, eight participants that did not follow study instructions, and 12 participants for classification test scores that were two or more standard deviations away from mean group performance. The final sample consisted of 145 participants (*blocked studying* group, $n = 37$; *interleaved studying* group, $n = 36$; *blocked pretesting* group, $n = 36$; *interleaved pretesting* group, $n = 36$), had a mean age of 29.5 years, and was 57% male. Forty-one percent of these participants were from the United Kingdom, 31% were from the United States, 13% were from Canada, 12% from Australia, and the remainder were from other countries; 28% of participants were Asian, 16% were Black, 4% were Mixed, 45% were White, and ~7% were from other ethnic groups or declined to provide ethnicity information.

Design, Materials, and Procedure

The design, materials, and procedure were based on the prior experiments, with a single session per participant (as in Experiment 1) and a prescreening conducted beforehand (as in Experiment 2). Unlike the prior experiments, however, time on task during training was equated across groups, with 80 s allotted per case study. Specifically, in both studying groups, participants spent 80 s per case study, whereas in both pretesting groups, participants had 40 s to guess the disorder and 40 s to read the answer. Pilot testing suggested that the 80 s duration was usually sufficient in all groups (i.e., allowing for learning to occur uninterrupted in all cases: blocking and interleaving; studying and pretesting). It could be argued, however, that this approach advantaged the studying groups by giving them more than ample time to examine the case studies in full knowledge of the presented disorder, whereas the pretesting groups still had to divide up the allotted time across multiple tasks, resulting in less time to study the correct answers.

Further, to address the first-trial accuracy disparity observed in the prior experiments, the interleaved groups used newly generated random sequences which ensured that no case study appeared more frequently in any trial position than any other. These patterns remained constrained such that no two case studies of the same disorder appeared consecutively.

Results

Results are presented in the same order and format as in the preceding experiments except that no analyses of time on task were needed.

Pretest Performance

Mean guessing performance for the blocked pretesting and interleaved pretesting groups is shown in the bottom panel of Figure 1. The same overall patterns as in the prior experiments were observed, except that there was no disparity in first-trial accuracy (as expected given the aforementioned modifications). The level of improvement was also more similar across than in prior experiments: In the blocked pretesting group, guessing accuracy improved from the first case study ($M = 0.39$, $SE = 0.027$) to the third case study ($M = 0.94$, $SE = 0.053$) of each disorder (~0.56 improvement), $t(35) = 16.10$, $p < .0001$, $d = 2.68$, and in the interleaved pretesting group, guessing accuracy improved from the first case study ($M = 0.41$, $SE = 0.035$) to the third case study ($M = 0.80$, $SE = 0.11$) of each disorder (~0.39 improvement), $t(34) = 8.73$, $p < .0001$, $d = 1.48$.

Classification Test Performance

Classification test results involving new and previously viewed case studies are displayed in the bottom panel of Figure 2.

New Case Studies. A 2 × 2 ANOVA with factors of Training Schedule (Blocked vs. Interleaved) and Training Activity (Studying vs. Pretesting) conducted on participant-level mean test scores for new case studies revealed a significant effect of Training Schedule, $F(1, 141) = 8.09$, $p = .0051$, $\eta_p^2 = 0.054$; a significant effect of Training Activity, $F(1, 141) = 9.47$, $p = .0025$, $\eta_p^2 = 0.063$; and no significant interaction ($p = .25$). These results, which replicate those of the preceding experiments, are consistent with patterns evident in Figure 2 (bottom panel, left side), in which there are indications of an advantage of interleaving over blocking, pretesting over studying, and the best overall performance in the interleaved pretesting group.

Previously Viewed Case Studies. A 2 × 2 ANOVA with factors of Training Schedule (Blocked vs. Interleaved) and Training Activity (Studying vs. Pretesting) conducted on participant-level mean test scores for previously viewed case studies revealed a significant effect of Training Schedule, $F(1, 141) = 7.30$, $p = .0078$, $\eta_p^2 = 0.049$; no significant effect of Training Activity, $F(1, 141) = 3.06$, $p = .082$, $\eta_p^2 = 0.021$; and no significant interaction ($p = .71$). Those results resemble patterns found in the prior experiments, except that an overall advantage of pretesting did not emerge. In an inspection of Figure 2 (bottom panel, right side), however, there were clear indications of an advantage of interleaving over blocking and best overall performance in the interleaved pretesting group. It is also notable that one third of participants in the interleaved pretesting group received full scores; as such, a ceiling effect in that group may have attenuated the pretesting advantage.

Metacognitive Questions

Data from the metacognitive questions (Table 2) were analyzed using 2 × 2 ANOVAs as in the preceding experiments. The following significant results were obtained; all other main effects or interactions not mentioned here were not significant. For judgments of learning, there were significant main effects of Training Schedule, $F(1, 141) = 4.54$, $p = .035$, $\eta_p^2 = 0.031$, and Training Activity, $F(1, 141) = 10.67$, $p = .0014$, $\eta_p^2 = 0.072$; these results reflect higher judgments in the interleaved and pretesting groups. For judgments of difficulty, there was a significant main effect of Training Activity, $F(1, 141) = 10.95$, $p = .0012$, $\eta_p^2 = 0.072$, reflecting higher difficulty ratings in the pretesting groups. For predictions of future test performance for new cases, there was a significant main effect of Training Activity, $F(1, 141) = 9.70$, $p = .0022$, $\eta_p^2 = 0.64$, as was observed for predictions of future test performance for previously viewed cases, $F(1, 141) = 25.32$, $p < .0001$, $\eta_p^2 = 0.15$. Those differences reflected higher predictions from participants in the pretesting versus studying groups, in line with patterns observed in the prior experiments.

Discussion

Across three experiments, substantial benefits of interleaving, pretesting, and their combination were observed for acquiring categorical knowledge and classification skills. Interleaving enhanced learning over blocking, replicating the interleaving effect for inductive learning of psychopathological disorders (e.g., Zulkiply et al., 2012). Moreover, pretesting yielded better learning than studying, establishing a pretesting effect for category learning and classification skills (see Figure 2). Additionally, the combination of interleaving and pretesting yielded the highest performance on both a 5-min (Experiments 1 and 3) and 48-hr delayed (Experiment 2) classification test. Finally, the pretesting advantage over studying was not solely due to differential time on task (Experiment 3). Overall, this study demonstrates that interleaving and pretesting are complementary learning strategies, with their combination improving category learning and classification skills more than either strategy alone.

How Interleaving and Pretesting Enhanced Learning

How did each learning strategy enhance category learning and classification skills? Regarding interleaving, the juxtaposition of different disorders in the interleaved groups likely enabled participants to focus on identifying the features that differentiate one disorder from another (Carvalho & Goldstone, 2019; S. H. K. Kang & Pashler, 2012; Zulkiply et al., 2012), yielding better understanding of defining characteristics. That process of discriminative contrast may have also been aided by temporal spacing between examples of each disorder, resulting in increased memory retrieval and better recall of relevant information on the classification test (S. H. K. Kang, 2017; Zulkiply et al., 2012). The net result was an improved ability to classify both brand-new and previously viewed case studies.

Regarding pretesting, attempts to answer pretest questions likely enhanced memory for case studies and the distinguishing characteristics of the disorders (for related theorizing, see Mera et al., 2022; Pan & Carpenter, 2023). Such attempts may have improved learning by prompting a search for identifying features, increasing curiosity, and encouraging closer attention. Further, correct answer feedback (in which participants compared their guess with the correct answer) provided opportunities to generate memories for the specific disorders and refine understanding through mental hypothesis testing (cf. Do & Thomas, 2023). Thus, although pretesting reduced the time for studying each case study (i.e., half or more of the time per case study was devoted to guessing the disorder represented by the case study), it was clearly more beneficial for developing classification skills.

As just described, pretesting likely facilitated learning via multiple pathways. Some of these pathways align with theories from basic category learning research, wherein guessing with feedback has been shown to be beneficial (e.g., Ashby et al., 2002; Carvalho & Goldstone, 2015), contrasting with theoretical accounts of the pretesting effect for paired associate materials (e.g., mediator generation). Additionally, the psychopathological disorders being studied were generally indistinguishable by simple rules. Instead, learners had to consider multiple characteristics together to determine category membership—a process that, in visual category learning, is known to benefit significantly from guessing with feedback (Ashby et al., 2002). Thus, the observed pretesting effects may share mechanisms with those hypothesized in other category learning research.

Training phase performance in the pretesting groups sheds light on the learning processes that occurred during interleaving and blocking. In the blocked pretesting groups, the oscillating accuracy pattern reflects the ease of guessing disorders in the second and third case studies of each block. Participants likely anticipated seeing additional examples of the same disorder, potentially delaying engagement with defining characteristics until prompted by an incorrect guess (for related discussion, see Rohrer & Pashler, 2010). Such patterns were unlikely in the interleaved pretesting group, wherein each successive case study involved a different disorder, enhancing the unpredictability that can bolster the effectiveness of interleaving (Pan et al., 2019). An analogous situation may have also occurred in the blocked studying versus interleaved studying groups.

Interleaved Pretesting and Interleaving Versus Pretesting

Regarding the combination of interleaving and pretesting, the present results enable us to adjudicate between the potential outcomes outlined at the outset of this article. First, it is evident that interleaving and pretesting do not necessarily engage cognitive mechanisms or affect learning in an antagonistic manner. Combining strategies did not yield deleterious effects. Second, interleaving and pretesting do not appear to yield learning processes that are highly redundant or nullify each other. If so, then interleaved pretesting probably would have performed on par with interleaved studying or blocked pretesting.

Ultimately, the present results suggest that the combination of interleaving and pretesting yields learning benefits that are *at least* additive (based on a comparison of interleaving and pretesting effect magnitudes in each experiment). Additive effects may result from two pathways: First, interleaving and pretesting may yield separate boosts to the same learning processes (e.g., identifying defining features); alternatively, the two strategies may enhance learning via different mechanisms (e.g., comparison of case studies; hypothesis testing). We suspect that the most likely scenario involved a mix of both pathways.

It is important to note, however, that the observed pretesting effects may have been attenuated by ceiling effects in the interleaved pretesting groups. For new case studies, 12.5%, 10%, and 19% of participants in that group achieved perfect scores in Experiments 1–3, respectively, whereas for previously viewed case studies, these proportions were 36%, 21%, and 36%, respectively. No other group had as many such participants near ceiling (the next highest being the blocked pretesting group where 22% reached ceiling for previously viewed case studies in Experiment 3). With greater room for improvement (i.e., if the classification test questions were somewhat more difficult), the pretesting effect would likely have been larger (including for old case studies in Experiment 3). If so, then the combination of interleaving and pretesting may have yielded synergistic benefits.[1]

From a metacognitive perspective, although the pretesting and studying groups did not consistently differ in difficulty ratings, pretesting led to higher judgments of learning (Experiments 2 and 3) and higher predictions of future test performance (all experiments). Those results potentially stem from participants drawing on their experiences of answering pretest questions, unlike in the studying groups (for related discussion, see Pan & Rivers, 2023). Such patterns also contrast with metacognitive findings for other types of materials (e.g., paired associates, visual materials), where participants often fail to recognize the advantages of more effective strategies. An intriguing possibility is that the cognitive mechanisms underlying the benefits of interleaved pretesting may differ from those in studies involving other materials, potentially making them more readily appreciated by learners.

Limitations and Directions for Future Research

Future research could explore whether the benefits of interleaving and pretesting generalize to other forms of category learning, to authentic educational settings, different retention intervals, and other types of skills. For example, pretesting could have involved being presented with the disorder name and guessing its defining characteristics. A reviewer helpfully pointed out that doing so would be a closer analogue to prior pretesting studies (e.g., Richland et al., 2009) than given characteristics and having to guess the category (as in the present study). Indeed, the type of pretesting examined in this study differs from that of many other pretesting studies in both the target materials and the types of guesses that participants were asked to make. It remains to be determined whether similar learning enhancements would be observed in situations where pretesting procedures are more reminiscent of those used, for example, with

[1] We thank Sean Kang for highlighting this possibility.

learning from expository texts. Moreover, we did not observe indications that the pretesting effect was larger at a longer retention interval (cf. Kliegl et al., 2024).

Other variations of interleaved pretesting (e.g., with different guessing procedures) could also be examined. Participants may have responded differently if, for example, more case studies were used (which for blocked pretesting would have increased participants' sense that the same disorder was being repeatedly shown), or the predictability of the presentation sequence differed. In other contexts (e.g., grammar learning), randomized interleaving can be more effective for learning than systematic interleaving (e.g., Pan et al., 2019). Studies using probabilistic sequences could also introduce unpredictability to both blocking and interleaving (e.g., Carvalho & Goldstone, 2015).

A further limitation of the present research is that participants were not informed about the upcoming classification test. The pretesting groups' guessing procedure also resembled the classification test, which raises the prospect of transfer-appropriate processing (i.e., where encoding information similar to its retrieval enhances recall performance; Morris et al., 1977). Future studies should investigate if pretesting benefits persist under conditions where test expectancy is stronger and subsequent test procedures are different.

This study represents at least the third instance wherein the combination of evidence-based study scheduling and learning activity strategies yields substantial learning benefits. Other examples include spacing and retrieval practice (Rawson & Dunlosky, 2022) and interleaving and retrieval practice (Sana & Yan, 2022). Future research could also help determine whether interleaving and pretesting always yield additive benefits and whether other such combinations are beneficial for learning.

Practical Implications

The present research reveals that the combination of interleaving and pretesting can be more effective for acquiring categorical knowledge and classification skills than either strategy alone. That result suggests that instructors and students can accrue substantial learning benefits from interleaved pretesting. To do so, instructors will need to provide pretesting opportunities followed by correct answer feedback, plus arrange learning in an interleaved manner. Students will need to be familiarized with interleaved pretesting, which typically entails many erroneous guesses and a more gradual rate of improvement during practice (as is a hallmark of many "desirable difficulties"; Bjork & Bjork, 2011). Improved category learning and better classification skills are the likely result.

References

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30*(5), 666–677. https://doi.org/10.3758/BF03196423

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

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, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 59–68). Worth Publishers.

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, *145*(11), 1029–1052. https://doi.org/10.1037/bul0000209

Carpenter, S. K., & Pan, S. C. (2024). Spacing effects in learning and memory. In L. Mickes & J. T. Wixted (Eds.), *Cognitive psychology of memory, Vol. 2 learning and memory: A comprehensive reference* (3rd ed.). Academic Press. https://doi.org/10.1016/B978-0-443-15754-7.00020-1

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, *1*, 496–511. https://doi.org/10.1038/s44159-022-00089-1

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, *22*(1), 281–288. https://doi.org/10.3758/s13423-014-0676-4

Carvalho, P. F., & Goldstone, R. L. (2019). When does interleaving practice improve learning? In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (1st ed., pp. 411–436). Cambridge University Press. https://doi.org/10.1017/9781108235631.017

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

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

Do, L. A., & Thomas, A. K. (2023). The underappreciated benefits of interleaving for category learning. *Journal of Intelligence*, *11*(8), Article 153. https://doi.org/10.3390/jintelligence11080153

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie* [On memory: Studies in experimental psychology]. Duncker & Humblot.

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

Foster, N. L., Mueller, M. L., Was, C., Rawson, K. A., & Dunlosky, J. (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Memory & Cognition*, *47*(6), 1088–1101. https://doi.org/10.3758/s13421-019-00918-4

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*(4), 505–513. https://doi.org/10.3758/s13421-011-0174-0

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*(4), 514–527. https://doi.org/10.3758/s13421-011-0167-z

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition*, *41*(5), 625–637. https://doi.org/10.3758/s13421-013-0298-5

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*(6), 1441–1451. https://doi.org/10.1037/a0020636

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. https://doi.org/10.1002/acp.1801

Kang, S. H. K. (2017). The benefits of interleaved practice for learning. In J. Horvath, J. Lodge, & J. Hattie (Eds.), *From the laboratory to the classroom: Translating science of learning for teachers* (pp. 79–93). Routledge.

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, *103*(1), 48–59. https://doi.org/10.1037/a0021977

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), Article 51. https://doi.org/10.1007/s10648-023-09772-y

Kliegl, O., Bartl, J., & Bäuml, K.-H. T. (2024). The pretesting effect comes to full fruition after prolonged retention interval. *Journal of Applied Research in Memory and Cognition*, *13*(1), 63–70. https://doi.org/10.1037/mac0000085

McDaniel, M. A. (2023). Combining retrieval practice with elaborative encoding: Complementary or redundant? *Educational Psychology Review*, *35*(3), Article 75. https://doi.org/10.1007/s10648-023-09784-8

Mera, Y., Rodriguez, G., & Marin-Garcia, E. (2022). Unraveling the benefits of experiencing errors during learning: Definition, modulating factors, and explanatory theories. *Psychonomic Bulletin & Review*, *29*(3), 753–765. https://doi.org/10.3758/s13423-021-02022-8

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Murphy, C. S. (2017). *Examining the boundaries of the spacing effect in inductive learning* [Unpublished dissertation]. University of Memphis.

Murphy, C. S., & Pavlik, P. I. (2018). Effects of spacing and testing on inductive learning. *Journal of Articles in Support of the Null Hypothesis*, *15*(1), 23–40.

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and pretesting effects: A review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review*, *35*(4), Article 97. https://doi.org/10.1007/s10648-023-09814-5

Pan, S. C., Lovelett, J. T., Phun, V., & Rickard, T. C. (2019). The synergistic benefits of systematic and random interleaving for second language grammar learning. *Journal of Applied Research in Memory and Cognition*, *8*(4), 450–462. https://doi.org/10.1016/j.jarmac.2019.07.004

Pan, S. C., & Rivers, M. L. (2023). Metacognitive awareness of the pretesting effect improves with self-regulation support. *Memory & Cognition*, *51*(6), 1461–1480. https://doi.org/10.3758/s13421-022-01392-1

Pan, S. C., Sana, F., Schmitt, A. G., & Bjork, E. L. (2020). Pretesting reduces mind wandering and enhances learning during online lectures. *Journal of Applied Research in Memory and Cognition*, *9*(4), 542–554. https://doi.org/10.1016/j.jarmac.2020.07.004

Pan, S. C., Yu, L. W., Hong, Y., Wong, M. J., Selverajan, G., & Kaku, M. (2024). *Individual differences in fluid intelligence moderate the interleaving effect for perceptual category learning* [Manuscript submitted for publication].

Permut, S., Fisher, M., & Oppenheimer, D. M. (2019). TaskMaster: A tool for determining when subjects are on task. *Advances in Methods and Practices in Psychological Science*, *2*(2), 188–196. https://doi.org/10.1177/2515245919838479

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283–302. https://doi.org/10.1037/a0023956

Rawson, K. A., & Dunlosky, J. (2022). Successive relearning: An underexplored but potent technique for obtaining and maintaining knowledge. *Current Directions in Psychological Science*, *31*(4), 362–368. https://doi.org/10.1177/09637214221100484

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15*(3), 243–257. https://doi.org/10.1037/a0016496

Roelle, J., Endres, T., Abel, R., Obergassel, N., Nückles, M., & Renkl, A. (2023). Happy together? On the relationship between research on retrieval practice and generative learning using the case of follow-up learning tasks. *Educational Psychology Review*, *35*(4), Article 102. https://doi.org/10.1007/s10648-023-09810-9

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*(3), 355–367. https://doi.org/10.1007/s10648-012-9201-3

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, *39*(5), 406–412. https://doi.org/10.3102/0013189X10374770

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, *3*(4), 241–249. https://doi.org/10.3102/00028312003004241

Sana, F., & Carpenter, S. K. (2023). Broader benefits of the pretesting effect: Placement matters. *Psychonomic Bulletin & Review*, *30*(5), 1908–1916. https://doi.org/10.3758/s13423-023-02274-6

Sana, F., & Yan, V. X. (2022). Interleaving retrieval practice promotes science learning. *Psychological Science*, *33*(5), 782–788. https://doi.org/10.1177/09567976211057507

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16–27. https://doi.org/10.3758/s13421-012-0238-9

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*(3), 215–221. https://doi.org/10.1016/j.learninstruc.2011.11.002

Appendix

Example Case Studies

Dysfunctional Cognition Disorder

Mel, age 14, has poor performance in school. Her teachers observe that she has difficulties with reading comprehension passages and answering simple questions. She is unable to correctly identify the relevant content in the passage and rephrase them to correctly answer the questions. Moreover, she is unable to solve math problems that require abstract thinking and does not understand simple jokes made by her peers during math class. Mel struggles to understand humor or metaphorical language. Mel finds it difficult to express herself appropriately and make friends in school. She also notes difficulties with independent living skills, such as managing personal hygiene, cooking, and managing finances.

Cyclothymic Affect Disorder

Jen, age 37, is a mother of two kids and is a business owner. She often experiences episodes of feeling elevated and increased energy and becomes very productive during this time period,

(*Appendix continues*)

taking on multiple tasks and completing them well. However, sometimes she also becomes very irritable, yells at her kids, and exaggerates small miscommunications with her husband, causing disturbances in her personal relationships. Additionally, Jen finds it difficult to sleep, and she often feels hopeless and extremely sad, thinking that she is a bad wife or mother. Her frequent fluctuations in mood and behavior affect both her household chores and her business work.

Note. The common names for dysfunctional cognition disorder and cyclothymic affect disorder are attention-deficit hyperactivity disorder and bipolar disorder, respectively.

Received February 20, 2024
Revision received July 10, 2024
Accepted July 15, 2024 ■