American Psychological Association

Journal of Experimental Psychology:
Learning, Memory, and Cognition

© 2018 American Psychological Association
0278-7393/19/$12.00

2019, Vol. 45, No. 2, 272–287
http://dx.doi.org/10.1037/xlm0000576

Sleep Reduces the Testing Effect—But Not After Corrective Feedback and Prolonged Retention Interval

Magdalena Abel, Valerie Haller, Hanna Köck, and Sarah Pötschke
Regensburg University

Dominik Heib and Manuel Schabus
University of Salzburg

Karl-Heinz T. Bäuml
Regensburg University

Retrieval practice relative to restudy of learned material typically attenuates time-dependent forgetting. A recent study examining this testing effect across 12-h delays filled with nocturnal sleep versus daytime wakefulness, however, showed that sleep directly following encoding benefited recall of restudied but not of retrieval practiced items, which reduced, and even eliminated, the testing effect after sleep (Bäuml, Holterman, & Abel, 2014). The present study investigated, in 4 experiments, whether this modulating role of sleep for the testing effect is influenced by two factors that have previously been shown to increase the testing effect: corrective feedback and prolonged retention intervals. Experiments 1a and 1b applied 12-h delays and showed benefits of sleep for recall after both restudy and retrieval practice with feedback, but not after retrieval practice without feedback. Experiments 2a and 2b applied 24-h or 7-day delays and failed to observe any long-lasting benefits of sleep directly after encoding, on both restudied and retrieval practiced items. These results indicate that both corrective feedback and prolonged retention intervals reduce the modulating role of sleep for the testing effect as it can be observed after 12-h delays and in the absence of corrective feedback, which suggests a fairly limited influence of sleep on the effect.

Keywords: memory consolidation, restudy, retrieval practice, sleep, testing effect

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Much to most people's disappointment, our memories fade as time passes. As a consequence, we are more likely to fail at recalling sought-after information, which cannot only be frustrating on a personal level, but can also have real negative consequences, for instance, in academic or other professional contexts. Some research suggests that forgetting across time may affect most memories equally, irrespective of how well they were initially learned (e.g., Bahrick, 1984; Slamecka & McElree, 1983), so it may seem that, with enough time, we will to a great proportion forget what we once knew. Yet, research also shows that there may be some exceptions to this rule. For instance, at least to a certain degree, time-dependent forgetting can be reduced by complementing learning with tests and retrieval practice, a finding known as the testing effect (e.g., Carrier & Pashler, 1992; for a review, see Roediger & Butler, 2011). Similarly, sleep in comparison with wakefulness after encoding has been suggested to stabilize memories and to decrease time-dependent forgetting (for a review, see Diekelmann & Born, 2010).

Effects of Retrieval Practice and Sleep on Time-Dependent Forgetting

In typical testing effect studies, subjects who completed an initial study phase are asked to practice the studied material by means of further study cycles (i.e., restudy) or retrieval-practice cycles (i.e., tests for the initially studied material). Final recall is assessed after both shorter retention interval (e.g., after 5 min) and more prolonged retention interval (e.g., after several days; e.g., Pyc & Rawson, 2010; Wheeler & Roediger, 1992). After shorter retention interval, restudy mostly results in similar or even higher recall rates compared with retrieval practice. Regardless of these initial recall levels, however, retrieval practice typically increases long-term retention with prolonged retention interval and reduces time-dependent forgetting relative to restudy, which results in significant test–delay interactions (e.g., Roediger & Karpicke, 2006; Wheeler, Ewers, & Buonanno, 2003). Direct or indirect testing effects may then arise, dependent on initial recall levels. If recall at short delay is rather similar between restudy and retrieval practice conditions, the reduction in time-dependent forgetting may create a direct testing effect, that is, significantly enhanced recall after retrieval practice compared with restudy after longer delay (e.g., Mulligan & Picklesimer, 2016; Toppino & Cohen,

This article was published Online First April 26, 2018.

Magdalena Abel, Valerie Haller, Hanna Köck, and Sarah Pötschke, Department of Experimental Psychology, Regensburg University; Dominik Heib and Manuel Schabus, Center for Cognitive Neuroscience, University of Salzburg; Karl-Heinz T. Bäuml, Department of Experimental Psychology, Regensburg University.

This work was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) awarded to Karl-Heinz T. Bäuml and Manuel Schabus (BA 1382/14-1).

Correspondence concerning this article should be addressed to Magdalena Abel, Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany. E-mail: magdalena.abel@ur.de

2009). Yet, if recall at short delay is superior in the restudy condition, recall after longer delay may lead to similar recall levels in retrieval practice and restudy conditions, thus showing an indirect testing effect as evidenced by a significant test–delay interaction and reduced time-dependent forgetting in response to retrieval practice (e.g., Hogan & Kintsch, 1971; Kornell, Bjork, & Garcia, 2011; Smith, Roediger, & Karpicke, 2013; Thompson, Wenger, & Bartling, 1978).

Prolonged retention intervals of several days do not just consist of time spent awake, but also comprise substantial periods of sleep. Critically, research on sleep-associated memory consolidation indicates that the type of delay interval that follows relatively closely upon learning can affect retention as well. Studies addressing the issue often compare two conditions with delay intervals of the same duration, which differ in whether they contain sleep or not. For instance, one frequently used approach is to ask subjects to study information either in the morning (e.g., at 9 a.m.) or in the evening (e.g., at 9 p.m.), and to come back for a memory test after a delay interval of roughly 12 hr. The typical finding is that subjects who studied in the evening and were tested after a night with sleep show enhanced recall relative to subjects who studied in the morning and were tested after a regular day filled with wakefulness (e.g., Fenn & Hambrick, 2013; Payne, Stickgold, Swanberg, & Kensinger, 2008; Scullin & McDaniel, 2010). The effect is often attributed to sleep-associated memory consolidation, assuming that memory contents are reactivated during certain sleep stages, which strengthens and stabilizes them (for reviews, see Rasch & Born, 2013; Stickgold, 2013). Although research on prolonged retention intervals is scarce, some studies reported sleep benefits also after longer delay than just 12 hr (e.g., Gais, Lucas, & Born, 2006; Griessenberger et al., 2012; Mazza et al., 2016; Stickgold, James, & Hobson, 2000; Wagner, Hallschmid, Rasch, & Born, 2006; but see Schönauer, Grätsch, & Gais, 2015), suggesting that sleep-associated memory consolidation can have a long-lasting influence on remembering.

The Role of Sleep for the Testing Effect

Building on the previous work on the testing effect on the one hand and sleep-associated memory consolidation on the other, a recent study by Bäuml et al. (2014) examined the interplay of the two effects. In this study, subjects were presented a set of study material, engaged in subsequent restudy or retrieval practice on the material, and were then asked to come back to the lab for a final memory test after 12 hr, which were either filled with nighttime sleep or daytime wakefulness. Within each of four experiments, results from short delay control conditions showed no major difference in recall between restudy and retrieval practice. In contrast, after longer delay, a direct testing effect with higher recall after retrieval practice compared with restudy emerged after 12 hr filled with wakefulness but not after 12 hr filled with sleep, leading to significant test–delay interactions (i.e., reduced time-dependent forgetting after retrieval practice) after the 12-h wake delay but not after the 12-h sleep delay. The pattern arose because sleep after encoding was beneficial for contents that had been subject to restudy, but left memories that had been subject to retrieval practice largely unaffected (for a related finding, see Abel & Bäuml, 2012). All of these findings accrued irrespective of which study materials were used (e.g., semantically categorized item lists, unrelated paired associates, prose passages), whether one or two practice cycles were applied, whether retroactive interference was induced before the final test or not, or whether recall levels were matched or unmatched across restudy and retrieval-practice conditions.

Bäuml et al. (2014) explained their findings in terms of the distribution-based bifurcation model of the testing effect (see Halamish & Bjork, 2011; Kornell et al., 2011; see Figure 1 for an illustration). According to this model, both restudy and retrieval practice increase memory strength of practiced items. Whereas restudy strengthens all items about equally, though to a moderate degree only (see Figure 1a), retrieval practice creates a bifurcated distribution of items: items not retrieved during practice remain at their original strength level and thus below recall threshold, whereas successfully retrieved items are strengthened to a rather high degree (see Figure 1b). With prolonged delay, the model assumes that all three item types (i.e., restudied, successfully retrieved, and nonretrieved items) decrease in strength at a comparable rate. Yet, successfully retrieved items may remain above recall threshold much longer than the restudied items that were strengthened to a lower degree, resulting in the emergence of a typical testing effect (i.e., higher recall after retrieval practice compared with restudy) and a test–delay interaction (i.e., less time-dependent forgetting after retrieval practice than restudy).

On the basis of this model, Bäuml et al. (2014) suggested that if sleep strengthened all three item types (i.e., restudied, successfully retrieved, and initially nonretrieved items) about equally, then sleep may reduce or even eliminate the testing effect and the test–delay interactions. Indeed, whereas sleep-associated memory consolidation may help restudied items to stay above recall threshold, resulting in a typical benefit of sleep for recall (see Figure 1a), both initially nonretrieved items and retrieved items may not show any sleep-associated recall benefits: initially nonretrieved items may fall too far below recall threshold with delay to be lifted above threshold through sleep-associated strengthening, but successfully retrieved items may still be above recall threshold after delay, not leaving much room for additional benefits of sleep-associated strengthening (see Figure 1b). Following this reasoning, benefits of sleep on recall should emerge mainly after restudy but not after retrieval practice, which should reduce or even eliminate the testing effect and any test–delay interactions. Bäuml et al.'s (2014) finding of reduced time-dependent forgetting and enhanced recall after retrieval practice compared with restudy after wake but not after sleep delay supports this rationale.

Two Open Research Questions

The study by Bäuml et al. (2014) made a first step in examining the interplay between sleep and the testing effect by investigating whether sleep can influence the testing effect when retrieval practice is conducted without feedback and when 12-h delay intervals between study and test are employed. The present study extends this prior work by addressing two further research questions on the interplay between sleep and the testing effect. The one question is whether feedback during retrieval practice can influence the effect of sleep on the testing effect; the other question is whether the results reported in Bäuml et al. generalize to longer retention intervals than 12 hr between study and test.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

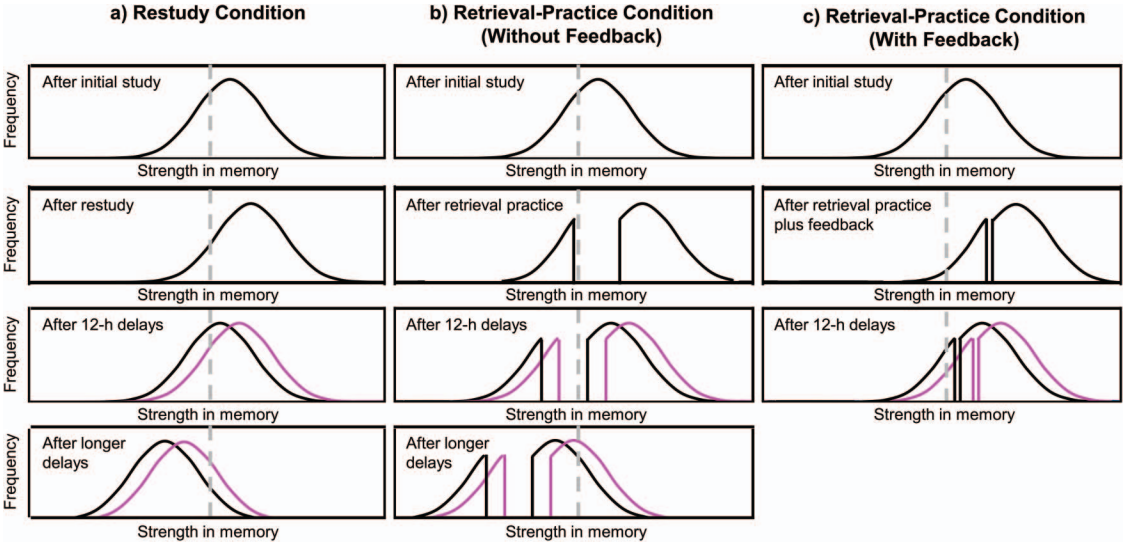

Figure 1. Illustration of memory strength distributions of three hypothetical sets of items, following Kornell, Bjork, and Garcia's (2011) bifurcation model. Column 1a shows items that were restudied, column 1b shows items that received retrieval practice (without corrective feedback), and column 1c shows items that received retrieval practice plus corrective feedback. The first horizontal line of panels shows memory strength after one initial study cycle; at this point, the three distributions are identical. The second horizontal line of panels shows how the distributions are changed through restudy and retrieval practice with and without feedback. Whereas restudy increases memory strength for all items, retrieval practice without feedback bifurcates the item distribution: successfully retrieved items are strengthened to a relatively high degree, but nonretrieved items remain at their initial strength level. In contrast, when corrective feedback is provided during retrieval practice, initially nonretrieved items show a pronounced benefit from feedback, which (partly) enables them to cross the recall threshold (vertical dashed line in gray). The third horizontal line of panels shows how distributions may be affected by 12-h delays, including either diurnal wakefulness (black curves) or nighttime sleep (pink [gray] curves). All items show the same amount of time-dependent forgetting and sleep-associated strengthening. Restudied items benefit from sleep and remain above recall threshold with a higher probability in the presence of sleep-associated strengthening. In contrast, benefits of sleep may be harder to detect after retrieval practice without feedback, because successfully retrieved items are still too high above recall threshold to show any additional benefits, whereas initially nonretrieved items may not be strengthened enough to cross recall threshold. If feedback is provided after retrieval practice, however, sleep-associated strengthening may help to keep (initially nonretrieved) items above recall threshold. The last horizontal line of panels shows how item distributions may be affected by even longer delays (e.g., delays of 24 hr or 7 days). Because all items are assumed to decrease in strength at a similar rate across delay, even items that were successfully retrieved during retrieval practice may cross below recall threshold when such prolonged retention intervals are applied, thereby unmasking potential additional benefits of sleep-associated strengthening. See the online article for the color version of this figure.

Bäuml et al. (2014) compared restudy cycles with pure retrieval practice cycles. Yet, considerable research on the testing effect indicates that the benefits of retrieval practice can be further enhanced when testing is combined with corrective feedback (e.g., Kang, McDermott, & Roediger, 2007; Pashler, Cepeda, Wixted, & Rohrer, 2005). Corrective feedback provides an additional learning opportunity so that initial mistakes can be corrected, uncertainties be reduced, and correct answers be maintained (e.g., Butler, Karpicke, & Roediger, 2008; Smith & Kimball, 2010; see also Roediger & Butler, 2011). In fact, in real life, learners seeking to memorize specific information are unlikely to rely on retrieval practice alone and are more likely to complement it with corrective feedback. Knowledge about whether corrective feedback changes the role of sleep for the testing effect is therefore of high empirical relevance.

In addition, following typical studies on the role of sleep for memory (e.g., Fenn & Hambrick, 2013; Scullin & McDaniel, 2010), Bäuml et al. (2014) focused on retention intervals of 12 hr, filled with either nighttime sleep or daytime wakefulness, to examine the interplay of sleep and the testing effect. The testing effect, however, has been shown to increase with prolonged retention intervals of up to several days (e.g., Roediger & Karpicke, 2006). Consistently, a recent meta analysis found larger testing effects in studies that applied retention intervals of at least one day compared with studies that applied shorter retention intervals (Rowland, 2014). Although some prior work on sleep-associated memory consolidation suggests that sleep effects may persist across more prolonged retention intervals than just a night of sleep or a day of wakefulness (e.g., Gais et al., 2006; Stickgold et al., 2000; Wagner et al., 2006), it is

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

unclear whether this holds true for the interplay of sleep and the testing effect as well. In daily life, students may often cram for an exam the next day, but education is focused on learning more generally, and certainly above and beyond 12-h retention intervals. It is therefore important to know whether prolonged delay changes the role of sleep for the testing effect.

The two research questions addressed in the present study are also of theoretical relevance, and expectations on the answers to these questions can be derived from the bifurcation model when the model is again enriched by the assumption that all item types benefit from sleep-induced strengthening and do so in a comparable way (see Bäuml et al., 2014). First, within this model, corrective feedback may reduce or even eliminate the bifurcated item distribution in retrieval-practice conditions, because mainly not successfully retrieved items should be subject to strengthening through feedback, whereas successfully retrieved items should hardly be affected by feedback, if at all (Kornell et al., 2011; Pashler et al., 2005; Pastötter & Bäuml, 2016). In such case, the nonretrieved items may be lifted above recall threshold and thus become susceptible to time-dependent forgetting, showing forgetting rates similar to those in a restudy condition. Critically, when corrective feedback brings initially nonretrieved items above recall threshold, then additional sleep-associated strengthening should be able to keep an even greater proportion of these items above recall threshold with delay, resulting in a beneficial effect of sleep on recall not only after restudy but also in the retrieval-practice condition with feedback (see Figure 1c). Test–delay interactions should therefore be eliminated, and time-dependent forgetting be similar after restudy and retrieval practice with feedback, irrespective of whether sleep or wakefulness follows upon encoding. Experiments 1a and 1b were conducted to address the issue and to test whether feedback changes the role of sleep for the testing effect.

Second, although the bifurcation model assumes that all contents show a similar decrease in strength with delay, the different degrees of strengthening of retrieved and restudied items may keep successfully retrieved items longer above recall threshold than items that were restudied, thus reducing time-dependent forgetting for retrieved items and creating the testing effect. However, if sleep strengthened items, and strengthened all items to a similar degree, then sleep should improve recall of restudied but not recall of successfully retrieved items after moderately long retention intervals (e.g., after 12 hr), thus reducing or even eliminating the testing effect (see Bäuml et al., 2014). In contrast, after severely prolonged retention intervals (e.g., after several days), additional sleep-associated strengthening of successfully retrieved items should improve also recall of the retrieved items, because also these items start to fall below recall threshold (Figure 1b). If so, the role of sleep for the testing effect and test–delay interactions would differ between moderate and long delay, with testing effect and test–delay interactions being absent after moderate delay, but being present after long delay. Importantly, such interactions in recall may arise although sleep may affect the strength of the single item types in a comparable, noninteractive way. Experiments 2a and 2b address the issue and examine whether prolonged delay can change the role of sleep for the testing effect.

Experiments 1a and 1b

The goal of Experiments 1a and 1b was to examine whether corrective feedback after retrieval practice can change the role of sleep for the testing effect. In both experiments, subjects studied unrelated paired associates and then engaged in restudy, retrieval practice without corrective feedback, and retrieval practice with corrective feedback for equal parts of the initially studied materials. A final memory test on all paired associates was conducted after a 12-h delay that included either nighttime sleep or daytime wakefulness; in addition, a short-delay control condition was included to (a) examine potential time-of-day effects and (b) assess test–delay interactions (i.e., time-dependent forgetting). One practice cycle was placed after initial encoding in Experiment 1a and three practice cycles were applied in Experiment 1b. Following the rationale above, we expected to replicate the results of Bäuml et al. (2014) and observe significant benefits of sleep after restudy, but not after retrieval practice in the absence of corrective feedback. This finding should arise in both experiments irrespective of initial practice level (see Bäuml et al., 2014). In particular, we expected benefits of sleep-associated strengthening to emerge after retrieval practice with corrective feedback as well. If so, differences in time-dependent forgetting and, as a consequence, differences in test–delay interactions between the two retrieval practice conditions should arise.

Experiment 1a

Method

Participants. Sample sizes in all present experiments were chosen so as to be similar to the sample sizes applied in Bäuml et al. (2014). Originally, 115 students from Regensburg University were recruited for Experiment 1a. Seven participants had to be excluded prior to data analysis because they either reported alcohol intake or daytime napping between sessions. A final sample of 108 healthy participants remained (M = 23.9 years; range 18–32 years; 43 male). Subjects were tested either individually or in pairs, and were distributed equally across conditions (n = 36 in each of the three delay conditions). For practical reasons, subject distribution could not be done via full random assignment. If subjects' personal schedules did not allow participation in a certain delay condition, we allowed participation in a different condition. This approach was chosen for all data collections reported in this article. Moreover, all experiments were conducted so as to be compatible with the declaration of Helsinki as adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964 and amended by the 64th WMA General Assembly, Fortaleza, Brazil, October 2013.

Materials. Materials consisted of 36 unrelated paired associates that were created by pairing two single items taken from different semantic categories (Scheithe & Bäuml, 1995; Van Overschelde, Rawson, & Dunlosky, 2004), using one as stimulus and one as response term. The material was randomly divided into three sets containing 12 paired associates each; across participants, these sets were equally often assigned to the three practice conditions and thus subject to restudy, retrieval practice without corrective feedback, and retrieval practice with corrective feedback.

Design. The experiment had a 3 × 3 mixed-factorial design with the between-subjects factor of delay (short delay control, 12-h

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

wake, 12-h sleep) and the within-subjects factor of type of practice (restudy, retrieval practice without feedback, retrieval practice with feedback). After initial study, all subjects were asked to engage in restudy for one third of the material, in retrieval practice without corrective feedback for another third, and in retrieval practice with corrective feedback for the last third. In the 12-h wake condition, participants studied and practiced the paired associates at 9 a.m., and the final test was conducted at 9 p.m., after 12 hr of wakefulness; in contrast, in the 12-h sleep condition, participants studied and practiced the paired associates at 9 p.m., and took the final test at 9 a.m., after one night of nocturnal sleep (for similar designs, see Abel & Bäuml, 2013; Bäuml et al., 2014; Payne et al., 2008; Scullin & McDaniel, 2010). Because learning and test sessions took place at different times of day across delay conditions, an additional short-delay condition was included to control for potential circadian effects. Half of the subjects in this condition participated at 9 a.m., the other half at 9 p.m., with a short delay of 5 min between learning phase and test. Apart from acting as a circadian control, the short-delay condition is a precondition to assess time-dependent forgetting across the 12-h delays.

Procedure.

Study and practice phase. In an initial study phase, all 36 paired associates were presented successively and in a random order, at a presentation rate of 5 sec per word pair. Subsequently, three practice blocks followed. On each block, one third of the initially studied paired associates was practiced. Blocks differed in which type of practice was carried out. On the block that provided a restudy opportunity, 12 of the paired associates were reexposed in intact form, in a random order and at an 8-s rate. On the block on which subjects were asked to engage in retrieval practice without corrective feedback, the stimulus terms of 12 paired associates plus the initial letters of the corresponding response terms were presented as retrieval cues for 8 sec each and in a random order. Subjects were asked to try to complement the presented cues with the correct response terms and to write their answers on a sheet of paper. On the block on which corrective feedback was presented after retrieval practice, stimulus terms of 12 paired associates plus the initial letters of the response terms were presented as cues in a random order as well, but subjects were only given 5 sec to complement the cues with the correct response terms. After 5 sec, the intact word pair was presented on the screen for 3 sec, thus providing corrective feedback. Subjects were instructed to complement the presented cues before feedback was provided; in addition, the experimenter was present during the whole experiment and ensured that subjects followed this instruction. All three practice blocks comprised only one practice opportunity for the three sets of materials (i.e., there were no further repetitions). Sequence of practice conditions as well as material in the three practice conditions were counterbalanced across participants.

The learning phase was followed by a distractor phase of 5 min, during which participants engaged in an unrelated cognitive distractor task. Afterward, subjects in the short-delay control conditions completed the final recall test. In contrast, subjects in the 12-h delay conditions were asked to leave the lab and to return to take the same final memory test after a delay of 12 hr that was either spent awake or filled with normal nighttime sleep. Subjects in the 12-h sleep conditions reported to have slept regularly during the night (M = 7.9 hrs; SD = 1.2), whereas subjects in the 12-h wake conditions reported not to have taken naps during the day. None of the subjects included in the final sample reported alcohol intake between sessions.

Test phase. At test, subjects were presented with all paired associates' stimulus terms plus the initial letter of the corresponding response terms and were asked to write down the correct response terms. Retrieval cues were presented in random order and for 8 sec each. When the final test was completed, subjects were debriefed and thanked for their participation.

Results

Success rates during retrieval-practice cycles. A 3 × 2 ANOVA with the factors of delay (short delay, 12-h wake delay, 12-h sleep delay) and retrieval practice (with and without corrective feedback) showed a marginally significant main effect of retrieval practice, reflecting a numerically higher success rate when no corrective feedback was provided and subjects had more time to engage in retrieval practice (64.3% vs. 60.7%), $F(1, 105) = 2.85$, $MSE = 238.72$, $p = .094$, $\eta^2 = .03$. No other effects were close to significance, all Fs ≤ 1.54, all ps $\geq .219$.

Circadian control (short-delay condition). Table 1 shows mean recall levels on the final test after the short delay. A 3 × 2 ANOVA with the factors of type of practice (restudy, retrieval practice without feedback, retrieval practice with feedback) and time of day (9 a.m., 9 p.m.) revealed a significant main effect of type of practice, $F(2, 68) = 28.73$, $MSE = 146.11$, $p < .001$, $\eta_p^2 = .46$. Recall was better after restudy than after retrieval practice with feedback (74.3% vs. 65.1%), $t(35) = 3.88$, $p < .001$, $d = 0.65$, and it was better after retrieval practice with feedback than after retrieval practice without feedback (65.1% vs. 52.8%) $t(35) = 4.37$, $p < .001$, $d = 0.73$. No other effects reached significance, indicating that recall was unaffected by circadian effects, all Fs < 1. For all further analyses we therefore collapsed the short-delay data, no longer differentiating between the time-of-day conditions.

Final test performance and time-dependent forgetting across delays. Figure 2 shows recall performance in the three delay conditions. A 3 × 3 ANOVA with the factors of type of practice (restudy, retrieval practice without feedback, retrieval practice with feedback) and delay (short delay, 12-h sleep, 12-h wake) revealed significant main effects of type of practice, $F(2, 210) = 23.52$, $MSE = 157.91$, $p < .001$, $\eta_p^2 = .18$, and delay, $F(2, 105) = 5.09$, $MSE = 1620.34$, $p = .008$, $\eta_p^2 = .09$. Moreover, there was a significant interaction between the two factors, $F(4, 210) =$

Table 1
Mean Recall (Plus Standard Deviations) in the Short-Delay Control Condition of Experiment 1a as a Function of Time of Day (9 a.m., 9 p.m.) and Practice Format (Restudy, Retrieval Practice Without Feedback, Retrieval Practice With Feedback)

Time	Restudy	Retrieval practice without feedback	Retrieval practice with feedback
9 a.m.	75.0 % (19.2)	53.7 % (17.0)	65.7 % (18.3)
9 p.m.	73.6 % (28.3)	51.9 % (19.9)	64.4 % (24.7)
Combined	74.3 % (23.9)	52.8 % (18.3)	65.1 % (21.4)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

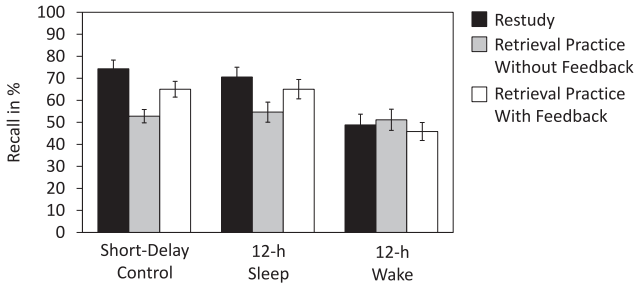

Figure 2. Mean recall in Experiment 1a as a function of delay (short-delay control, 12-h sleep, 12-h wake) and practice format (restudy, retrieval practice without feedback, retrieval practice with feedback). Error bars represent ±1 standard errors.

9.84, $MSE = 157.91$, $p < .001$, $\eta_p^2 = .16$, indicating that practice type modulated how the different delay intervals affected memory performance. To further examine test–delay interactions, we calculated separate 2 × 2 ANOVAs comparing forgetting in retrieval-practice conditions (with and without corrective feedback) relative to restudy conditions across delays.

ANOVAs that contrasted restudy and retrieval practice without feedback revealed a significant test–delay interaction from short delay across the 12-h wake delay, $F(1, 70) = 29.81$, $MSE = 171.60$, $p < .001$, $\eta_p^2 = .30$, whereas no significant test–delay interaction was observed across the 12-h sleep delay, $F(1, 70) = 1.73$, $MSE = 160.46$, $p = .193$, $\eta_p^2 = .02$. In contrast, ANOVAs that contrasted restudy and retrieval practice with feedback failed to reveal any significant test–delay interactions, both after the 12-h wake delay, $F(1, 70) = 2.09$, $MSE = 168.59$, $p = .153$, $\eta_p^2 = .03$, and the 12-h sleep delay, $F(1, 70) = 1.06$, $MSE = 116.84$, $p = .307$, $\eta_p^2 = .02$.

Follow-up t tests showed that there was indeed no significant time-dependent forgetting in any of the practice conditions across the 12-h sleep delay, all ts(70) < 1. Both after the short delay and the sleep delay, restudy led to higher recall than retrieval practice with feedback, and retrieval practice with feedback led to higher recall than retrieval practice without feedback, all ts(35) ≥ 2.06, all ps ≤.047, all ds ≥ 0.34. In contrast, there were roughly equivalent amounts of time-dependent forgetting across the 12-h wake delay after restudy and retrieval practice with feedback, both ts(70) ≥ 3.19, ps = .002, ds ≥ 0.75, whereas virtually no time-dependent forgetting became evident after retrieval practice without feedback, $t(70) < 1$. After the 12-h wake delay, recall no longer differed between the three practice types, $F(2, 70) = 1.24$, $MSE = 206.30$, $p = .295$, $\eta_p^2 = .03$.

Consistent with the above analyses, a final 3 × 2 ANOVA contrasting the 12-h wake and sleep delays showed not only significant main effects for type of practice, $F(2, 140) = 5.17$, $MSE = 165.90$, $p = .007$, $\eta_p^2 = .07$, and delay, $F(1, 70) = 6.27$, $MSE = 1891.53$, $p = .015$, $\eta_p^2 = .08$, but also a significant interaction, $F(2, 140) = 10.64$, $MSE = 165.90$, $p < .001$, $\eta_p^2 = .13$, reflecting benefits of sleep for some, but not all practice levels. Sleep compared with wakefulness was beneficial for restudied contents (70.6% vs. 48.8%), $t(70) = 3.30$, $p = .002$, $d = 0.78$, and also for contents that had been subject to retrieval practice plus corrective feedback (65.1% vs. 45.8%), $t(70) = 3.19$, $p = .002$, $d = 0.75$. Contents that had been subject to retrieval practice without corrective feedback were largely unaffected by sleep, however (54.6% vs. 51.2%), $t(70) < 1$.

Discussion

The results of Experiment 1a demonstrate that corrective feedback can change the role of sleep for the testing effect. A significant test–delay interaction reflecting reduced time-dependent forgetting after retrieval practice was observed when restudy was compared with retrieval practice without feedback across a 12-h delay filled with wakefulness, but not across a 12-h delay filled with sleep, which replicates Bäuml et al. (2014). Yet, when restudy was compared with retrieval practice with feedback, no corresponding test–delay interaction was observed. Time-dependent forgetting across the 12-h wake delay affected restudied items and items subject to retrieval practice with feedback to a similar degree, whereas items subject to retrieval practice without feedback showed no such forgetting. Similarly, sleep in comparison to wakefulness was beneficial for memories after restudy and retrieval practice with feedback, but not after retrieval practice without feedback. All of these results indicate that corrective feedback after retrieval practice can change the role of sleep for the testing effect.

Experiment 1a showed an indirect testing effect, but no direct testing effect (i.e., better recall after retrieval practice compared with restudy after the 12-h delay), which differs from the Bäuml et al. (2014) study. This difference in results between studies very likely emerged because differences in recall levels between restudy and retrieval practice after short delay were much larger in the present than in the prior study. In the present experiment, after short delay, restudy enhanced recall relative to the retrieval practice condition by roughly 20%, whereas Bäuml et al. (2014) reported no or much smaller corresponding benefits for restudy in their short-delay control conditions. As is indicated from prior work (e.g., Hogan & Kintsch, 1971; Kornell et al., 2011; Smith et al., 2013), differences in initial recall levels typically influence the presence of a direct testing effect, whereas they do not affect the presence of test–delay interactions, which is consistent with the present results.

Experiment 1b

Bäuml et al. (2014) showed that, without corrective feedback after retrieval practice, the role of sleep for the testing effect did not depend on practice level. Sleep was beneficial for restudied, but not retrieval practiced contents, irrespective of whether one or two practice cycles were carried out. Experiment 1b was conducted as a conceptual replication of Experiment 1a, but a second goal was to examine whether sleep's role for the testing effect would also be unaffected by practice level in the presence of corrective feedback, not just in its absence. Therefore, three practice cycles were conducted after initial study in Experiment 1b (instead of one as in Experiment 1a). On the basis of the results from the prior work, we expected practice level to not affect the role of sleep for the testing effect.

Method

Participants. One hundred fourteen students from Regensburg University were recruited for the experiment. Six participants

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

had to be excluded prior to data analysis because of reported alcohol intake or daytime napping. A final sample of 108 participants remained (M = 21.9 years; range 18–30 years; 18 male). Subjects were distributed equally across conditions (n = 36 in each condition).

Materials. Materials comprised 36 new unrelated paired associates that were created by pairing two single items taken from different semantic categories (Scheithe & Bäuml, 1995; Van Overschelde et al., 2004). As in Experiment 1a, the material was randomly divided into three sets containing 12 paired associates each; sets were equally often assigned to the three practice conditions.

Design. The experiment had the same 3 × 3 mixed-factorial design as Experiment 1a. The factor delay (short delay control, 12-h wake, 12-h sleep) was again manipulated between subjects, the factor type of practice (restudy, retrieval practice without feedback, retrieval practice with feedback) was manipulated within subjects.

Procedure. The procedure was identical to that of Experiment 1a, with only one exception. The three practice blocks that followed upon initial study and differed in whether paired associates were restudied or retrieval practiced (with or without corrective feedback) now each comprised three practice cycles. On each block, the respective paired associates were practiced in a random order; when the first practice cycle was complete, paired associates were practiced again, in a new random order. Each block was completed with a final third practice cycle, again conducted in a new random order. All other procedural details were identical to Experiment 1a.

Results

Success rates during retrieval-practice cycles. A 3 × 3 × 2 ANOVA with the factors of delay (short delay, 12-h wake delay, 12-h sleep delay), practice cycle (first, second, third) and retrieval practice (with and without corrective feedback) revealed a significant main effect of retrieval practice, $F(1, 105) = 30.25$, $MSE = 661.09$, $p < .001$, $\eta_p^2 = .22$. Across practice cycles, retrieval success was higher when corrective feedback was provided (88.3% vs. 77.2%). There was also a significant main effect of practice cycle, $F(2, 210) = 157.98$, $MSE = 160.19$, $p < .001$, $\eta_p^2 = .60$, indicating that success rates improved across practice cycles (73.0% vs. 86.3% vs. 88.8%). Whereas no other effects reached significance, all Fs < 1, there was a significant interaction between retrieval practice and practice cycle, $F(2, 210) = 99.54$, $MSE = 185.05$, $p < .001$, $\eta_p^2 = .49$. Although success rates increased across practice cycles even when no corrective feedback was provided (75.7% vs. 77.3% vs. 78.5%), ts(107) ≥ 2.67, ps ≤.009, ds ≥.26, the improvement was greater when corrective feedback was present (70.3% vs. 95.3% vs. 99.2%, ts(107) ≥ 4.74, ps <.001, ds ≥.46).

Circadian control (short-delay condition). Table 2 shows mean recall levels on the final test after the short delay. A 3 × 2 ANOVA with the factors of type of practice (restudy, retrieval practice without feedback, retrieval practice with feedback) and time of day (9 a.m., 9 p.m.) showed a significant main effect of type of practice, $F(2, 68) = 16.36$, $MSE = 299.28$, $p < .001$, $\eta_p^2 = .33$. There was no difference in recall after restudy and retrieval practice with feedback (92.7% vs. 93.4%), $t(35) < 1$, but both practice types led to better recall than retrieval practice without feedback (75.3%), ts(35) ≥ 4.23, ps <.001, ds ≥ 0.71. Even though there were numerical differences between a.m. and p.m. control conditions (see Table 2 for details), no other effects reached significance, indicating that recall was largely unaffected by circadian effects, all Fs < 1. For all further analyses we again collapsed the short-delay data and no longer differentiated between the time-of-day conditions.

Table 2
Mean Recall (Plus Standard Deviations) in the Short-Delay Control Condition of Experiment 1b as a Function of Time of Day (9 a.m., 9 p.m.) and Practice Format (Restudy, Retrieval Practice Without Feedback, Retrieval Practice With Feedback)

Time	Restudy	Retrieval practice without feedback	Retrieval practice with feedback
9 a.m.	90.3 % (15.2)	75.7 % (28.3)	90.3 % (16.4)
9 p.m.	95.1 % (8.7)	75.0 % (23.9)	96.5 % (7.2)
Combined	92.7 % (12.5)	75.4 % (25.8)	93.4 % (12.9)

Final test performance and time-dependent forgetting across delays. Figure 3 shows recall performance in the three delay conditions. A 3 × 3 ANOVA with the factors of type of practice (restudy, retrieval practice without feedback, retrieval practice with feedback) and delay (short delay, 12-h sleep, 12-h wake) revealed significant main effects of type of practice, $F(2, 210) = 25.85$, $MSE = 201.50$, $p < .001$, $\eta_p^2 = .20$, and delay, $F(2, 105) = 5.59$, $MSE = 986.43$, $p = .005$, $\eta_p^2 = .10$. In addition, there was a significant interaction between the two factors, $F(4, 210) = 3.07$, $MSE = 201.50$, $p = .017$, $\eta_p^2 = .06$, indicating that practice type modulated how the different delay intervals affected recall. As before, to further examine test–delay interactions, we conducted separate 2 × 2 ANOVAs comparing forgetting in retrieval-practice conditions (with and without corrective feedback) relative to restudy conditions across delays.

ANOVAs comparing restudy and retrieval practice without feedback revealed a significant test–delay interaction from short delay across the 12-h wake delay, $F(1, 70) = 9.65$, $MSE = 188.96$, $p = .003$, $\eta_p^2 = .12$, but not across the 12-h sleep delay, $F(1, 70) < 1$. In contrast, ANOVAs comparing restudy and retrieval practice with feedback failed to reveal any significant test–delay interactions, both after the 12-h wake delay and the 12-h sleep delay, Fs(1, 70) < 1.0.

Follow-up t tests further confirmed this pattern. There was no significant time-dependent forgetting in any of the practice conditions across the 12-h sleep delay, all ts(70) ≤ 1.54, all ps ≥.128, all ds ≤ 0.36. After both short delay and 12-h sleep delay, recall did not differ between restudy and retrieval practice with feedback, but both practice types led to higher recall than retrieval practice without feedback, ts(35) ≥ 4.23, all ps ≤.001, all ds ≥ 0.59. In contrast, there were roughly equivalent amounts of time-dependent forgetting across the 12-h wake delay after restudy and retrieval practice with feedback, ts(70) ≥ 4.23, ps <.001, ds ≥ 1.00, whereas no significant time-dependent forgetting arose after retrieval practice without feedback, $t(70) < 1$. After the 12-h wake delay, recall no longer differed between practice types, $F(2, 70) = 1.08$, $MSE = 235.68$, $p = .337$, $\eta_p^2 = .03$.

Consistently, a final 3 × 2 ANOVA contrasting the 12-h wake and sleep delays showed not only significant main effects for type

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

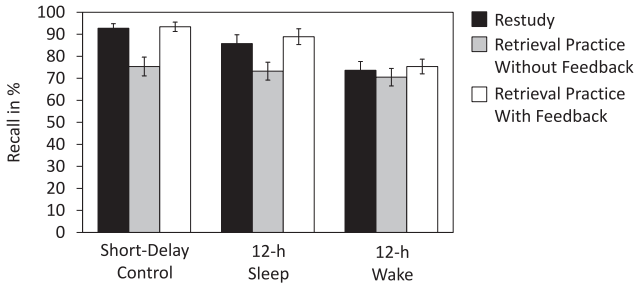

Figure 3. Mean recall in Experiment 1b as a function of delay (short-delay control, 12-h sleep, 12-h wake) and practice format (restudy, retrieval practice without feedback, retrieval practice with feedback). Error bars represent ±1 standard errors.

of practice, $F(2, 140) = 10.93$, $MSE = 188.66$, $p < .001$, $\eta_p^2 = .14$, and delay, $F(1, 70) = 4.01$, $MSE = 1214.20$, $p = .049$, $\eta_p^2 = .05$, but also a significant interaction, $F(2, 140) = 3.27$, $MSE = 188.66$, $p = .041$, $\eta_p^2 = .05$, reflecting benefits of sleep for only some practice conditions. *T* tests showed that sleep compared with wakefulness was again beneficial for restudied contents (85.8% vs. 73.6%), $t(70) = 2.15$, $p = .035$, $d = 0.51$, and contents that had been subject to retrieval practice plus corrective feedback (88.9% vs. 75.4%), $t(70) = 2.76$, $p = .007$, $d = 0.65$, whereas contents that had been subject to retrieval practice without corrective feedback did not benefit from sleep (73.3% vs. 70.5%), $t(70) < 1$.

Discussion

A comparison of results between Experiments 1a and 1b confirms that the role of sleep for the testing effect is not affected by practice level, irrespective of whether corrective feedback is provided after retrieval practice or not. Even after three practice cycles instead of one, the results in Experiment 1b were the same as those in Experiment 1a. In the absence of corrective feedback, sleep eliminated the test–delay interaction that was found across the 12-h wake delay; in the presence of corrective feedback, there was no test–delay interaction at all, both after the 12-h wake and the 12-h sleep delay. Moreover, as in Experiment 1a, the testing effect became evident in a significant test–delay interaction across the 12-h wake delay, but did not result in enhanced recall after retrieval practice compared with restudy. Like in Experiment 1a, this is very likely attributable to restudy leading to a pronounced (roughly 17%) benefit compared with retrieval practice after short delay (see also above). The findings of Experiment 1b thus mirror those of Experiment 1a in several ways.

On a theoretical level, the results of the two experiments are consistent with the bifurcation model when the model is enriched by the assumption that sleep strengthens the single item types in a comparable way. Indeed, by bringing initially nonretrieved items closer to recall threshold, according to the model, corrective feedback may not only eliminate the test–delay interaction observed when no feedback is provided, but may also enable initially nonretrieved items to benefit in recall from sleep-associated strengthening (see Figure 1c). Corrective feedback thus does not only change the testing effect itself, it can also set a limit to the modulating role of sleep for the effect.

Experiments 2a and 2b

The goal of Experiments 2a and 2b was to examine whether not only corrective feedback but also very long retention intervals between study and test can influence the role of sleep for the testing effect. Following typical prior sleep studies, Bäuml et al. (2014) employed 12-h delay intervals to investigate the role of sleep for the testing effect and observed that, in the absence of corrective feedback, testing effects and test–delay interactions arose after a 12-h wake delay but not after a 12-h sleep delay. This prior finding is consistent with the enriched bifurcation model (see above).

Yet, on the basis of the very same model, the prediction arises that the finding may not generalize across even longer retention intervals, but that regular testing effects and test–delay interactions may again be present when the delay is further increased, irrespective of whether sleep or wakefulness followed directly upon encoding. Because retrieval practice alone strengthens memories already to a rather high degree, benefits of sleep for successfully retrieved items may not affect recall after 12-h delays (see Figure 1, third line of panels). However, when the memory strength of items decreases with delay, even items that were successfully retrieved during practice may come closer to recall threshold with retention intervals of several days. Under such conditions, additional sleep-associated strengthening may become relevant and affect recall such that testing effects and test–delay interactions are again observed (see Figure 1, last line of panels). Experiments 2a and 2b address the issue by examining testing effects in the absence of corrective feedback after delay intervals of 12 hr (Experiment 2a), 7 days (Experiments 2a and 2b), and 24 hr (Experiment 2b).

Experiment 2a

Method

Participants. One hundred sixty-eight students from Regensburg University completed the experiment, but 8 participants had to be excluded prior to data analysis because of reported alcohol intake or daytime napping. This resulted in a final sample of 160 participants ($M = 22.3$ years, $SD = 2.9$; 34 male). Subjects were distributed equally across conditions ($n = 32$ in each of five delay conditions).

Materials. Thirty-two unrelated Finnish-German vocabulary pairs were selected from an online dictionary (https://defi.dict.cc). The vocabulary pairs were randomly divided into two sets of 16 pairs. Across subjects, each set was equally often assigned to the restudy and retrieval-practice conditions.

Design. The experiment had a 2 × 5 mixed-factorial design. The factor type of practice (restudy, retrieval practice) was varied within subjects, whereas the factor delay (short delay control, 12-h wake, 12-h sleep, 7-day wake, 7-day sleep) was manipulated between subjects. After initial study, subjects were asked to engage in restudy for one half of the material, but in retrieval practice (without feedback) for the other half. Encoding took place at either 9 a.m. or 9 p.m., but the retention interval placed before the final test differed across delay conditions. In the short-delay condition, subjects completed the test after 5 min. In the 12-h delay conditions, subjects returned to the lab to take the same test after 12 hr

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

that included either nighttime sleep or daytime wakefulness. In the 7-day delay conditions, subjects received the same instructions as subjects in the 12-h delay conditions, and were either asked to stay awake during the first 12 hr after encoding or to sleep regularly during the night. In these 7-day conditions, however, subjects were asked to return only after a week to complete the final test. Please note that the labels of the 7-day "wake" and 7-day "sleep" conditions only refer to the manipulation of type of activity right after encoding; during the rest of both 7-day delay intervals, all subjects followed their regular sleep-wake cycles, so that subjects in both conditions can be assumed to have spent more or less similar amounts of time awake and asleep within the course of the week. The study did not entail experimental sleep deprivation in any of the delay conditions (i.e., subjects in the 7-day wake condition were of course not asked to stay awake for 7 days).

Procedure.

Study and practice phase. Initially, subjects were asked to study Finnish-German vocabulary pairs for a later memory test. During study, vocabulary pairs were presented one at a time, in a random order, and for 6 sec each. Subsequently, there were two practice blocks, one of them comprising restudy cycles and the other one comprising retrieval-practice cycles; sequence of practice conditions was counterbalanced. On the restudy block, subjects were shown half of the initially studied vocabularies in intact form and were asked to make use of the additional study time. There were two restudy cycles, and on each cycle vocabulary pairs were presented in a new random order, for 6 sec each. On the retrieval-practice block, subjects were presented retrieval cues for the other half of the initially studied vocabularies. On each of the two practice cycles, the Finnish words were presented together with the German word stem, in random order and for 6 sec each. Subjects were asked to try to recall the German meaning of the vocabularies and to write their answers on a sheet of paper.

After practice, subjects were asked to work on an unrelated cognitive task for 5 min. Subjects in the short-delay condition then took the final test, whereas subjects in the other delay conditions left the lab and returned after either 12 hr or 7 days to take the same test. Subjects in these long-delay conditions who had completed encoding at 9 p.m. reported to have slept regularly during the night ($M = 7.8$ hrs; $SD = 1.1$); subjects who had completed encoding at 9 a.m. reported not to have taken naps during the subsequent day.

Test phase. At test, the Finnish word and the initial letter of the German meaning were presented for each vocabulary pair; order was set to random and cues were shown for 8 sec each. Subjects were asked to try to recall as many of the vocabulary pairs as possible and write down the German meaning of the words.

Results

Success rates during retrieval-practice cycles. A 5 × 2 ANOVA with the factors of delay (short delay, 12-h wake delay, 12-h sleep delay, 7-day wake delay, 7-day sleep delay) and practice cycle (first, second) revealed a significant main effect of practice cycle, $F(1, 155) = 37.31$, $MSE = 16.50$, $p < .001$, $\eta_p^2 = .19$, indicating that success rates improved from the first to the second retrieval-practice cycle (64.7% vs. 67.4%). No other effects were significant, all Fs < 1.18, all ps $\geq .323$, showing that success rates did not differ between delay conditions.

Circadian control (short-delay condition). Table 3 shows mean recall levels on the final test after the short delay. A 2 × 2 ANOVA with the factors of type of practice (restudy, retrieval practice) and time of day (9 a.m., 9 p.m.) showed a significant main effect of type of practice, $F(1, 30) = 25.61$, $MSE = 188.76$, $p < .001$, $\eta_p^2 = .46$, reflecting better recall after restudy than retrieval practice after the short delay (83.2% vs. 65.8%). Again, despite numerical differences between a.m. and p.m. control conditions (see Table 3), no other effects were significant, suggesting that recall was unaffected by circadian effects, all Fs ≤ 1.01, ps $\geq .323$. For all further analyses we again collapsed the 9 a.m. and 9 p.m. data to one short-delay control condition.

Final test performance and time-dependent forgetting across delays. Figure 4 shows recall performance in all delay conditions. A 2 × 5 ANOVA with the factors of type of practice (restudy, retrieval practice) and delay (short delay control, 12-h wake, 12-h sleep, 7-day wake, 7-day sleep) showed a significant main effect for type of practice, $F(1, 155) = 9.57$, $MSE = 183.60$, $p = .002$, $\eta_p^2 = .06$, a significant main effect for delay, $F(4, 155) = 32.16$, $MSE = 623.04$, $p < .001$, $\eta_p^2 = .45$, and a significant interaction of the two factors, $F(4, 155) = 7.95$, $MSE = 183.60$, $p < .001$, $\eta_p^2 = .17$, suggesting that how the delays affected recall depended on type of practice.

We first examined how recall was affected by the 12-h delays and again performed separate 2 × 2 ANOVAs to assess time-dependent forgetting from short delay across the two 12-h delays. Most importantly, a significant test–delay interaction emerged only across the 12-h wake delay, $F(1, 62) = 15.08$, $MSE = 132.77$, $p < .001$, $\eta_p^2 = .20$, but not across the sleep delay, $F(1, 62) = 1.13$, $MSE = 211.75$, $p = .292$, $\eta_p^2 = .02$. The pattern of results after the 12-h sleep delay was again similar to that in the short-delay condition, with restudy resulting in higher recall than retrieval practice, all ts$(31) \geq 3.08$, ps $\leq .004$, ds ≥ 0.54. In contrast, after the 12-h wake delay, recall was no longer different for restudied and retrieval practiced contents, $t(31) < 1$. Consistently, a 2 × 2 ANOVA comparing the 12-h sleep and wake delays showed a significant main effect for type of practice, $F(1, 62) = 9.03$, $MSE = 160.89$, $p = .004$, $\eta_p^2 = .13$, and a significant interaction, $F(1, 62) = 5.33$, $MSE = 160.89$, $p = .024$, $\eta_p^2 = .08$, but no significant main effect of delay, $F(1, 62) = 1.92$, $MSE = 993.21$, $p = .171$, $\eta_p^2 = .03$. T tests showed that sleep compared with wakefulness was again beneficial for restudied contents (71.7% vs. 58.8%), $t(62) = 2.17$, $p = .034$, $d = 0.54$, but not for contents that had been subject to retrieval practice (59.8% vs. 57.2%), $t(62) < 1$.

Concerning the prolonged retention intervals, two separate 2 × 2 ANOVAs compared recall in the two 7-day delay conditions to

Table 3
Mean Recall (Plus Standard Deviations) in the Short-Delay Control Condition of Experiment 2a as a Function of Time of Day (9 a.m., 9 p.m.) and Practice Format (Restudy, Retrieval Practice Without Feedback)

Time	Restudy	Retrieval practice without feedback
9 a.m.	85.2 % (14.4)	62.5 % (20.0)
9 p.m.	81.3 % (18.5)	69.1 % (17.5)
Combined	83.2 % (16.5)	65.8 % (18.8)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

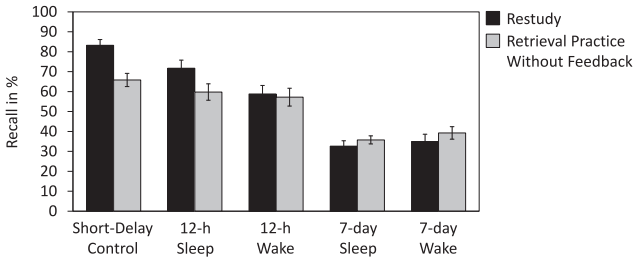

Figure 4. Mean recall in Experiment 2a as a function of delay (short-delay control, 12-h sleep, 12-h wake, 7-day sleep, 7-day wake) and practice format (restudy, retrieval practice without feedback). Error bars represent ±1 standard errors.

recall in the short-delay control condition and revealed significant test–delay interactions, all Fs(1, 62) ≥ 17.76, MSEs ≥ 189.48, ps < .001, $\eta_p^2 \geq .22$, indicating that retrieval practice compared with restudy reduced time-dependent forgetting across both 7-day delays and irrespective of whether sleep or wakefulness had followed upon initial encoding. Indeed, a final 2 × 2 ANOVA contrasting the 7-day wake and sleep delays showed no significant effects, all Fs(1,62) ≤ 2.14, MSEs ≤ 344.44, ps ≥ .149, $\eta_p^2 \leq .03$. After 7 days, recall was comparable after restudy and retrieval practice (33.8% vs. 37.5%), and it was not affected by whether subjects had slept or stayed awake right after encoding (34.2% vs. 37.1%).

Discussion

The results in the 12-h delay conditions replicate the results observed for retrieval practice without corrective feedback in Experiments 1a and 1b. Significant test–delay interactions reflecting reduced time-dependent forgetting after retrieval practice arose after 12 hr filled with wakefulness, but were absent after 12 hr filled with sleep. Going beyond Experiments 1a and 1b, Experiment 2a also shows that this specific effect of sleep does not generalize to retention intervals of 7 days. When recall was assessed after 7 days, significant test–delay interactions emerged irrespective of whether sleep or wakefulness had followed upon encoding.

On the basis of the bifurcation model and the assumption that sleep strengthens all types of memories in a comparable way, we had expected sleep effects on recall of both restudied and retrieval practiced contents after the 7-day delay. However, no evidence for long-lasting benefits of sleep-associated memory consolidation on recall after 7 days arose. The observed absence of a sleep effect on recall of restudied items is particularly surprising, because these items showed such sleep effect after the 12-h delay, indicating that the strengthening effect of sleep for these items disappeared with increasing delay. Our original reasoning was that the modulating role of sleep for the testing effect may be restricted to shorter delay, because sleep effects on recall would emerge irrespective of practice format after prolonged delay. Instead, the results suggest that the modulating role of sleep for the testing effect is restricted to shorter delay, because sleep improves recall of restudied items after 12-h delays, but does no longer do so after longer delay intervals.

Experiment 2b

The motivation for Experiment 2b was twofold. The first goal was to replicate the new finding of Experiment 2a, namely the nonpersistent effect of sleep on recall after prolonged delays of 7 days. For this, we applied the same type of stimulus material as employed in Experiments 1a and 1b (i.e., paired associates) to test whether beneficial effects of sleep on recall are transient with prolonged delay. The second goal was to evaluate how quickly sleep benefits on recall may fade. We therefore included both 7-day and 24-h delay conditions in the experiment. In both delay conditions, sleep and wake groups had similar amounts of sleep and wakefulness, the only difference being whether subjects, directly upon encoding, slept at night (and then stayed awake) or stayed awake during the day (and then went to sleep). We included the 24-h delay conditions, because they are among the shortest possible delay intervals with roughly equated times of sleep and wakefulness. We did not include 12-h delay conditions in this experiment, because these conditions were already part of Experiments 1a, 1b, and 2a. We did also not include additional short-delay conditions, because the focus of Experiment 2b was not on test–delay interactions. The results of the experiment will show whether the findings in the 7-day conditions of Experiment 2a can be replicated and whether they generalize to 24-h delay intervals.

Method

Participants. One hundred thirty-six students completed the experiment, but, again, 8 participants had to be excluded prior to data analysis because of reported alcohol intake or daytime napping. The final sample thus consisted of 128 participants (M = 22.6 years, SD = 2.5; 24 male) that were distributed equally across conditions (n = 32 in each of the four delay conditions).

Materials. Twenty-four unrelated paired associates were constructed by pairing items from different semantic categories (Scheithe & Bäuml, 1995; Van Overschelde et al., 2004). The material was randomly divided into two sets of 12 paired associates. Across subjects, each set was equally often assigned to the restudy and retrieval-practice conditions.

Design. The experiment had a 2 × 4 mixed-factorial design. The factor type of practice (restudy, retrieval practice) was again varied within subjects, the factor delay (24-h wake, 24-h sleep, 7-day wake, 7-day sleep) was manipulated between subjects. Subjects practiced paired associates by means of restudy and retrieval practice (without feedback) at either 9 a.m. or 9 p.m. In both 24-h delay conditions, subjects completed the final memory test after 24 hr: in the 24-h sleep condition, this delay interval included a period of sleep followed by a period of wakefulness; in the 24-h wake condition, the interval included a period of wakefulness followed by a period of sleep. In the 7-day delay conditions, subjects were also instructed to stay awake during the first 12 hr after encoding or to sleep regularly during the night. As in Experiment 2a, subjects in the 7-day delay conditions followed their regular sleep-wake cycles for the rest of the time and returned after 7 days to complete the final test.

Procedure.

Study and practice phase. Subjects were asked to memorize paired associates for a later memory test. Initially, word pairs were presented one at a time, in random order, and for 5 sec each. As in

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Experiment 2a, there were two subsequent practice blocks, one of them comprising restudy cycles, the other one comprising retrieval-practice cycles; sequence of practice conditions was counterbalanced. On the restudy block, subjects were reexposed to one half of the initially studied paired associates. There were two restudy cycles, and on each cycle word pairs were shown in a new random order, for 7 sec each. On the retrieval-practice block, subjects were presented with retrieval cues for the other half of the initially studied material. On each of the two practice cycles, the stimulus terms together with the initial letters of the corresponding response terms were presented in random order and for 7 sec each. Subjects were asked to complement the response terms and to write their answers on a sheet of paper. Afterward, subjects were asked to work on an unrelated cognitive task for 5 min before leaving the lab. Subjects returned after either 24 hr or 7 days to take the final memory test on all studied contents. Subjects who had completed encoding at 9 p.m. reported to have slept regularly during the subsequent night ($M = 8.2$ hrs; $SD = 0.9$); subjects who had completed encoding at 9 a.m. reported not to have taken naps during the subsequent day.

Test phase. At test, the stimulus term and the initial letter of the response term were presented for each word pair. Order was set to random and cues were shown for 7 sec each. Subjects were asked to write down the correct response terms.

Results

Success rates during retrieval-practice cycles. A 4 × 2 ANOVA with the factors of delay (24-h wake delay, 24-h sleep delay, 7-day wake delay, 7-day sleep delay) and practice cycle (first, second) revealed no significant effects, all Fs ≤ 2.39, all ps ≥ .125. Mean success rate was 86.7% on the first and 87.1% on the second practice cycle. It was unaffected by delay condition.

Final test performance. Figure 5 shows recall performance on the final test. A 4 × 2 ANOVA with the factors of delay (24-h wake delay, 24-h sleep delay, 7-day wake delay, 7-day sleep delay) and type of practice (restudy, retrieval practice) revealed a significant main effect of delay, $F(3, 124) = 17.22$, $MSE = 1027.84$, $p < .001$, $\eta_p^2 = .29$, reflecting better recall after the 24-h delays (54.5%) than after the 7-day delays (31.8%). No other effects were significant, all Fs(1, 124) ≤ 1.

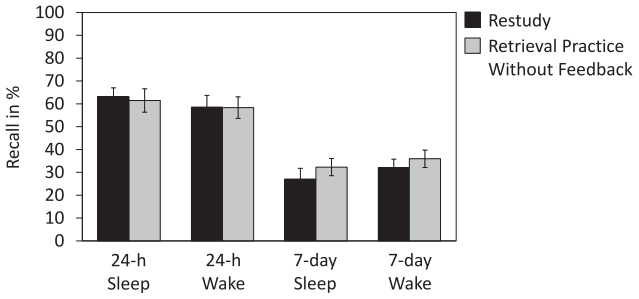

Figure 5. Mean recall in Experiment 2b as a function of delay (24-h sleep, 24-h wake, 7-day sleep, 7-day wake) and practice format (restudy, retrieval practice without feedback). Error bars represent ±1 standard errors.

Discussion

Applying the same type of study material as in Experiments 1a and 1b, Experiment 2b shows the same pattern of results as the long-delay recall data of Experiment 2a: Across prolonged delays of 7 days, there was no difference in recall between retrieval practice and restudy, and no effect of sleep on recall performance. The results of Experiment 2b also show that these findings generalize to the shorter 24-h conditions. Experiment 2b thus indicates that the moderating effect of sleep on the testing effect observed after 12-h delay intervals (see Experiments 1a, 1b, and 2a above) may be rather short-lived and already disappear after 24 hr.

Additional Analysis

Experiments 1a, 1b, and 2a all provided evidence in favor of testing effects, yet mostly indirectly, via significant test–delay interactions and reduced time-dependent forgetting. Indeed, in none of these experiments, a direct testing effect and enhanced recall after retrieval practice compared with restudy was observed, which may be because, in all three experiments, restudy clearly exceeded retrieval practice in the short delay condition (see above). All this holds while in the 7-day delay conditions of Experiments 2a and 2b, results showed a small trend for a direct testing effect. This trend did not turn out to be significant in the single experiments, but we examined whether the effect would be present when statistical power was enhanced, that is, when data of the 7-day delay conditions of Experiments 2a and 2b were pooled. Indeed, results of a 2 × 2 × 2 ANOVA with the factors of delay (7-day wake delay, 7-day sleep delay), type of practice (restudy, retrieval practice), and experiment (Experiment 2a, Experiment 2b) revealed a significant main effect of type of practice, $F(1, 124) = 4.74$, $MSE = 230.70$, $p = .031$, $\eta_p^2 = .04$, reflecting better recall after retrieval practice compared with restudy (34.1% vs. 29.6%), that is, a direct testing effect. No other effects were significant in this analysis, all Fs(1,124) ≤ 1.99, ps ≥ .161, $\eta_p^2 \leq .02$. The finding of no significant difference in recall between wake and sleep conditions indicates that the lack of a long-lasting effect of sleep in Experiments 2a and 2b was not due to low statistical power.

General Discussion

The starting point of the present set of experiments was a previous study by Bäuml et al. (2014), which had reported that 12-h delays filled with sleep can reduce or even eliminate the testing effect by benefiting recall of restudied items but leaving recall of items that had been subject to retrieval practice largely unaffected. In the present study, three of four experiments (Experiments 1a, 1b, and 2a) replicated this pattern: 12-h delays filled with sleep were again beneficial for recall of restudied items but not for recall of retrieval practiced items, thereby eliminating the test–delay interactions that were present across similar delays filled with daytime wakefulness. The new and more important finding of this study is that sleep after encoding does not always modulate the testing effect, and that this modulation is no longer present if corrective feedback is provided during retrieval practice and if prolonged retention intervals are applied.

Indeed, when retrieval practice was complemented by corrective feedback in Experiments 1a and 1b, the results showed largely

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

equivalent benefits of sleep for recall of restudied and retrieval practiced information after delays of 12-h. In contrast to corresponding analyses on retrieval practice without feedback, there were no significant test–delay interactions when comparing retrieval practice with corrective feedback to restudy, irrespective of whether the 12-h delay intervals were filled with nighttime sleep or daytime wakefulness; this finding arose both when one practice cycle (Experiment 1a) and when three practice cycles (Experiment 1b) were applied. Thus, corrective feedback cannot only change the testing effect and the consequences of retrieval practice for time-dependent forgetting (see Kornell et al., 2011), it also seems to set a limit to the modulating role of sleep for the testing effect.

Similarly, Experiments 2a and 2b showed that prolonged retention intervals of 7 days also reduce, or even eliminate, the influence of sleep on the testing effect as evident in test–delay interactions. Comparing retrieval practice without corrective feedback and restudy, significant test–delay interactions and reduced time-dependent forgetting after retrieval practice emerged after prolonged delays of 7 days, irrespective of whether sleep or wakefulness had followed upon encoding, and irrespective of whether vocabulary pairs (Experiment 2a) or unrelated paired associates (Experiment 2b) had been encoded. Moreover, sleep did not seem to have a persisting benefit for recall of any of the encoded contents across such prolonged delay. An additional 24-h delay condition included in Experiment 2b showed a similar pattern, suggesting that sleep benefits on recall may be quite short-lived and not extend from 12-h to 24-h delays. Prolonged retention intervals thus seem to set another limit to the modulating role of sleep for the testing effect.

Relation of the Present Findings to the Bifurcation Model

According to the bifurcation model, restudy and retrieval practice without corrective feedback both strengthen memories, but whereas restudy strengthens all restudied contents to about the same moderate degree, retrieval practice strengthens successfully retrieved items to a very high degree while leaving nonretrieved items unaffected—thereby bifurcating the item distribution. Assuming that sleep-associated strengthening affects all types of memories to a similar degree, Bäuml et al. (2014) argued that successfully retrieved items are already too far above recall threshold to show any additional sleep benefit after a delay interval of 12 hr, whereas nonretrieved items never crossed above recall threshold and, after delay, may have fallen too far below it to benefit from sleep. In contrast, the unbifurcated distribution of restudied items, with some items above and some items below recall threshold, enables restudied items to show an additional benefit, so that sleep-associated strengthening can keep a bigger proportion of them above threshold after a 12-h delay (see Figure 1). Both the present results on 12-h sleep and wake delays and retrieval practice without corrective feedback and the prior findings by Bäuml et al. (2014) show the expected pattern of results and thus are in line with the bifurcation model.

Experiments 1a and 1b examined the role of corrective feedback, and consistent with prior work on the testing effect, found that corrective feedback increased recall in retrieval practice conditions (e.g., Butler et al., 2008; Pashler et al., 2005) and eliminated test–delay interactions and differences in time-dependent forgetting relative to a restudy condition (e.g., Abel & Roediger, 2017; Kornell et al., 2011). Following the bifurcation model, corrective feedback is assumed to lift initially nonretrieved items above recall threshold (e.g., Kornell et al., 2011; Pastötter & Bäuml, 2016), so that these items can also cross below threshold with delay, thus contributing to time-dependent forgetting and reducing the test–delay interaction that is observed in the absence of corrective feedback. Similarly, because nonretrieved items are lifted above recall threshold by corrective feedback, recall of these items can show additional benefits from sleep-associated strengthening with delay, creating a sleep effect not only for the recall of restudied items but also for the recall of items that were subject to retrieval practice with corrective feedback. As a result, sleep may no longer modulate the testing effect. The results reported in Experiments 1a and 1b on the effects of corrective feedback show exactly such a pattern and thus are in line with the enriched bifurcation model.[1]

Experiment 2a examined prolonged retention intervals of 7 days and found significant test–delay interactions and reduced time-dependent forgetting after retrieval practice (without corrective feedback) irrespective of whether sleep or wakefulness had followed upon encoding. Because, in general, items decrease in memory strength with delay, even in the presence of additional sleep-associated strengthening, successfully retrieved items should cross below recall threshold after sufficiently long retention interval. In such case, the bifurcation model predicts that test–delay interactions should be intact as long as there is a bifurcation and as long as retrieval practice strengthens successfully retrieved items more than restudy strengthens the restudied items (see Kornell et al., 2011). The present results are consistent with this prediction, showing such test–delay interactions after both 7-day wake and 7-day sleep delay (and after both 24-h wake and 24-h sleep delay conditions; see below). However, on the basis of the additional assumption that sleep strengthens all types of memories, strengthens them to a comparable degree, and strengthens them persistently, we had expected that (a) the sleep effects on recall of restudied items in the 12-h delay conditions generalize to longer delay conditions, and (b) also sleep effects on recall of retrieval practiced items show up after longer delay. The results of Experiments 2a and 2b did not show such a pattern, indicating that the sleep effects in the present experiments were less persistent than expected (for more detailed discussion of this point, see below).

The existing literature on the testing effect reports both direct and indirect testing effects. Direct testing effects often show a pattern of similar recall of retrieval practice and restudy after short delay but improved recall after retrieval practice relative

[1] Another strength-threshold model that may also be in line with the present findings is Rickard and Pan's (in press) dual memory model. This model assumes that study and retrieval practice create separate memories: study memory and test memory, which consists of memory for the cue and the association between cue and response. Besides, the model includes the proposal that there is no learning on incorrect retrieval practice trials in the absence of feedback, thus incorporating the bifurcation phenomenon, and it claims that there is no bifurcation in the presence of feedback. The model also predicts a period of increasing testing effect magnitude as the delay between practice and test increases, as was observed in the present Experiment 2a.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

to restudy after longer delay (e.g., Mulligan & Picklesimer, 2016; Roediger & Karpicke, 2006; Toppino & Cohen, 2009). In contrast, indirect testing effects often show a pattern of higher recall after restudy than retrieval practice after short delay but similar recall between the two practice formats after longer delay (e.g., Kornell et al., 2011; Smith et al., 2013; Thompson et al., 1978). The results of the present study fall into this second category. Indeed, in all experiments reported in this study, restudy caused superior recall relative to retrieval practice after a short delay of few minutes (with recall advantages of roughly 15–20% in favor of restudy) but showed more time-dependent forgetting than retrieval practice, leading to similar levels of recall for the two practice conditions after longer delay. Only the pooled data of the 7-day delay conditions of Experiments 2a and 2b revealed a direct testing effect, with slightly enhanced recall after retrieval practice compared with restudy. The bifurcation model is consistent with both direct and indirect testing effects (see Kornell et al., 2011).

Relation of the Present Findings to the Elaboration and Episodic-Context Accounts of the Testing Effect

Although the bifurcation model fits well with the results of the present study, it is important to keep in mind that the bifurcation model is not a process model and is therefore silent on which cognitive processes underlie the testing effect (see Halamish & Bjork, 2011). Several accounts of the testing effect have suggested such cognitive mechanisms. Unfortunately, however, in themselves, these accounts make no direct predictions concerning the role of sleep for the effect, and further, may not easily enable predictions to be derived from related work on sleep-associated memory consolidation. For instance, the elaborative retrieval hypothesis (Carpenter, 2009, 2011; Pyc & Rawson, 2010) attributes the testing effect to the additional activation of semantically related information that is assumed to occur during retrieval practice, but not (or to a lesser degree) during restudy. Whereas some prior studies indicate that semantic associations related to studied information may show pronounced benefits from sleep (e.g., Cai, Mednick, Harrison, Kanday, & Mednick, 2009; McKeon, Pace-Schott, & Spencer, 2012; Payne et al., 2009), other recent studies point in a somewhat different direction (e.g., Chatburn, Kohler, Payne, & Drummond, 2017; Fenn, Gallo, Margoliash, Roediger, & Nusbaum, 2009; Landmann et al., 2016). On the basis of the elaborative retrieval hypothesis, it is therefore difficult to extract predictions regarding the role of sleep for the testing effect.

The case is similar for the episodic context account of the testing effect (Karpicke, Lehman, & Aue, 2014), which assumes that retrieval practice reactivates and updates episodic context associations that can later function as more effective retrieval cues during recall. Whereas some prior work indicates that sleep may decontextualize memories and make episodic context less important (Cairney, Durrant, Musgrove, & Lewis, 2011), other recent work found no such evidence for decontextualization (Cox, Tijdens, Meeter, Sweegers, & Talamini, 2014; Jurewicz, Cordi, Staudigl, & Rasch, 2016), again making it difficult to predict how sleep should affect the testing effect if changes in episodic context associations were the underlying cognitive process. Although current cognitive accounts of the testing effect thus may not lead to clear-cut predictions on the role of sleep for the testing effect, the results of the present study may serve as important empirical restrictions for future versions of these accounts, or completely new accounts of the testing effect.

How Persistent Are Sleep Benefits for Recall Performance?

Many studies show that intervals between study and test that are filled with sleep are beneficial for memory compared with intervals that are filled with wakefulness. This holds, although there is not an abundance of studies examining how long-lasting such sleep benefits really are. A few studies investigated the issue and reported sleep benefits across delays that go beyond 12-h intervals (Gais et al., 2006; Griessenberger et al., 2012; Mazza et al., 2016; Stickgold et al., 2000; Talamini, Nieuwenhuis, Takashima, & Jensen, 2008; Wagner et al., 2006), which is why initially we expected long-lasting sleep effects in our experiments, too (see above). The results of Experiments 2a and 2b of the present study show a different picture, however. Whereas benefits of sleep emerged consistently across 12-h delays, no such benefits arose after 7-day and 24-h retention intervals. Although this finding contrasts with the above studies, it is in line with the results of another recent study. In this study, Schönauer et al. (2015) reported sleep benefits for a declarative memory task (i.e., paired-associate learning) after 12-h delays, but failed to find such effects after delays of 3 and 6 days. The issue of whether sleep benefits are persistent across time thus may be more complex than is suggested from the existing literature, and, at least under certain circumstances, sleep benefits may vanish across delay intervals that go beyond 12 hr and just one night of sleep versus one day of wakefulness.

The present finding of transient benefits of sleep is also of theoretical relevance. Indeed, there are contrasting views on how sleep-associated benefits for memory arise (for a discussion, see Ellenbogen, Payne, & Stickgold, 2006). The classic perspective is that sleep passively protects memories from extraexperimental interference that accrues during wakefulness (e.g., Jenkins & Dallenbach, 1924; Wixted, 2004). In contrast, the currently prevailing perspective assumes an active role of sleep in memory consolidation, with memory contents being reactivated during certain sleep stages, thereby being stabilized and consolidated (e.g., Diekelmann & Born, 2010; Rasch, Büchel, Gais, & Born, 2007). The previous finding of persisting benefits of sleep following closely upon encoding, for instance, after 24-h intervals (e.g., Stickgold et al., 2000; Wagner et al., 2006) has been interpreted as evidence against a merely passive role of sleep (e.g., Gais et al., 2006; Talamini et al., 2008). Because such prolonged retention intervals contain similar amounts of sleep and wakefulness irrespective of whether sleep or wakefulness is placed after encoding, persisting benefits of sleep cannot easily be attributed to differing amounts of extraexperimental interference in sleep and wake conditions.

In contrast, both the present findings and those reported in Schönauer et al. (2015) may be interpreted as evidence that sleep shields memories from interference and, at least under certain circumstances, does little more for them (e.g.,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

strengthen and stabilize memories), so that sleep benefits become evident after shorter delay intervals that consist of sleep *or* wake delay, but do not arise after longer delay intervals that are matched for overall time spent awake and asleep.[2] With regard to the present study, this view may be further supported by the pattern of results for the 12-h delay conditions; whereas the results in the 12-h sleep conditions looked almost like exact copies of the results in the short-delay control conditions, the most interesting differences seemed to emerge across 12-h delays filled with wakefulness. The inconsistency in findings concerning the longevity of sleep benefits may therefore indicate that there is no simple answer to the question of how exactly sleep benefits memories, but that a perspective embracing contributions of sleep to both passive interference reduction and active consolidation might be most promising. Within such perspective, the relative contribution of sleep's passive versus active contribution may vary with experimental task, and sleep predominantly shelter memories from interference in some types of task, but predominantly strengthen and stabilize memories in other types of task. All this must remain speculative at this point in time and be investigated in more detail in future work.

A Possible Limitation of the Present Results

A caveat with regard to the present findings may be that, in all experiments, a cued-recall format was applied for both retrieval practice and test, in which additional letter cues were provided for the to-be-recalled response items. First, usage of letter cues could have fundamentally changed the nature of memory search on test trials, from purely episodic retrieval to partially semantic search from the letter cue followed by a check of the retrieved answer using episodic recognition memory. Such usage may also have made the retrieval task less difficult and thus may have reduced the size of the testing effects observed in this study, because testing effects have been shown to be larger when retrieval practice is more difficult (e.g., Carpenter & DeLosh, 2006; Kang et al., 2007; see also Rowland, 2014). Second, this procedural choice could have reduced ecological validity of our study, because learners in realistic scenarios may be unlikely to receive initial-letter cues during practice or test.

There is reason to expect that the core findings of this study would not have changed if no initial letter cues had been provided at practice and test. For instance, Bäuml et al. (2014) reported evidence that the role of sleep for the testing effect does not change when recall levels are varied, which suggests that harder retrieval practice (and reduced recall success) may still produce findings similar to the present ones. Also, in other contexts, the presence versus absence of initial letter cues during retrieval was found to have no major influence on how experimental factors influenced recall performance (e.g., Bäuml & Aslan, 2006; Bäuml & Samenieh, 2010), indicating that the absence of initial letter cues in the present study would not have led to qualitatively different results. Using settings with higher ecological validity than was employed in the present study, future work may like to address the issue, examining in more detail how different formats of retrieval practice and test influence the interplay of sleep and the testing effect.

Conclusions

This study shows that sleep can reduce, or even eliminate, the testing effect when retrieval practice is without corrective feedback and retention intervals are shorter than 24 hr. In contrast, when feedback is provided during retrieval practice or retention intervals of at least 24 hr are employed, typical testing effects arise. These results suggest that, in the long run, learners can apply retrieval practice and reap the resulting benefits irrespective of the timing of subsequent sleep or wakefulness phases. Under realistic study conditions, supplementing retrieval practice with corrective feedback, but engaging in continued practice even after initial retrieval success, has been suggested to be one of the most promising strategies for learners to boost long-term retention (e.g., Roediger & Butler, 2011). The present indication of a limited role of sleep for the testing effect supports this proposal.

[2] There are quite specific proposals in the literature on how exactly sleep may benefit memories. For instance, the synaptic homeostasis hypothesis assumes that sleep improves signal-to-noise ratio by down regulating global synaptic strengths (e.g., Tononi & Cirelli, 2014), selective tagging hypotheses assume that memories which received a molecular tag during encoding are preferentially consolidated (e.g., Payne & Kensinger, 2018), and network integration accounts assume that initially hippocampus-dependent memories are redistributed to cortical sites for long-term storage, which may allow integration with pre-existing memories (e.g., Born & Wilhelm, 2012). If refraining from pure speculation, none of these more specific and mostly neurobiologically grounded hypotheses allows predictions with regard to testing effects, or could explain why sleep effects emerge after 12 hrs, but are no longer present after longer retention intervals.

References

Abel, M., & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting, delay, and sleep. *Memory, 20,* 420–428.

Abel, M., & Bäuml, K.-H. T. (2013). Sleep can eliminate list-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 946–952.

Abel, M., & Roediger, H. L. (2017). Comparing the testing effect under blocked and mixed practice: The mnemonic benefits of retrieval practice are not affected by practice format. *Memory & Cognition, 45,* 81–92.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learning in school. *Journal of Experimental Psychology: General, 113,* 1–29.

Bäuml, K.-H. T., & Aslan, A. (2006). Part-list cuing can be transient and lasting: The role of encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 33–43.

Bäuml, K.-H. T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect - it enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1568–1581.

Bäuml, K.-H. T., & Samenieh, A. (2010). The two faces of memory retrieval. *Psychological Science, 21,* 793–795.

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research, 76,* 192–203.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 918–928.

Cai, D. J., Mednick, S. A., Harrison, E. M., Kanday, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 10130–10134.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Cairney, S. A., Durrant, S. J., Musgrove, H., & Lewis, P. A. (2011). Sleep and environmental context: Interactive effects for memory. *Experimental Brain Research, 214,* 83–92.

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

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552.

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

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642.

Chatburn, A., Kohler, M. J., Payne, J. D., & Drummond, S. P. A. (2017). The effects of sleep restriction and sleep deprivation in producing false memories. *Neurobiology of Learning and Memory, 137,* 107–113.

Cox, R., Tijdens, R. R., Meeter, M. M., Sweegers, C. C. G., & Talamini, L. M. (2014). Time, not sleep, unbinds contexts from item memory. *PLoS ONE, 9,* e88307.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11,* 114–126.

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: Passive, permissive, active or none? *Current Opinion in Neurobiology, 16,* 716–722.

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., III, & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16,* 509–513.

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review, 20,* 501–506.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory, 13,* 259–262.

Griessenberger, H., Hoedlmoser, K., Heib, D. P. J., Lechinger, W., Klimesch, W., & Schabus, M. (2012). Consolidation of temporal order in episodic memories. *Biological Psychology, 91,* 150–155.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 801–812.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562–567.

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology, 25,* 605–612.

Jurewicz, K., Cordi, M. J., Staudigl, T., & Rasch, B. (2016). No evidence for memory decontextualization across one night of sleep. *Frontiers in Human Neuroscience, 10,* 7.

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

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

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65,* 85–97.

Landmann, N., Kuhn, M., Maier, J. G., Feige, B., Spiegelhalder, K., Riemann, D., & Nissen, C. (2016). Sleep strengthens but does not reorganize memory traces in a verbal creativity task. *Sleep, 39,* 705–713.

Mazza, S., Gerbier, E., Gustin, M.-P., Kasikci, Z., Koenig, O., Toppino, T. C., & Magnin, M. (2016). Relearn faster and retain longer: Along with practice, sleep makes perfect. *Psychological Science, 27,* 1321–1330.

McKeon, S., Pace-Schott, E. F., & Spencer, R. M. C. (2012). Interaction of sleep and emotional content on the production of false memories. *PLoS ONE, 7,* e49353.

Mulligan, N. W., & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 938–950.

Pashler, H., Cepeda, N. J., Wixted, J., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3–8.

Pastötter, B., & Bäuml, K.-H. T. (2016). Reversing the testing effect by feedback: Behavioral and electrophysiological evidence. *Cognitive, Affective, and Behavioral Neuroscience, 16,* 473–488.

Payne, J. D., & Kensinger, E. A. (2018). Stress, sleep, and the selective consolidation of emotional memories. *Current Opinion in Behavioral Sciences, 19,* 36–43.

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L.-W., Wamsley, E. J., Tucker, M. A., . . . Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92,* 327–334.

Payne, J. D., Stickgold, R., Swanberg, K., & Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychological Science, 19,* 781–788.

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

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93,* 681–766.

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science, 315,* 1426–1429.

Rickard, T. C. & Pan, S. C (in press). A dual memory theory of the testing effect. *Psychonomic Bulletin & Review*.

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Rowland, C. A. (2014). The effect testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432–1463.

Scheithe, K., & Bäuml, K.-H. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien [German language norms for representatives of 48 categories]. *Sprache & Kognition, 14,* 39–43.

Schönauer, M., Grätsch, M., & Gais, S. (2015). Evidence for two distinct sleep-related long-term memory consolidation processes. *Cortex, 63,* 68–78.

Scullin, M. K., & McDaniel, M. A. (2010). Remembering to execute a goal: Sleep on it! *Psychological Science, 21,* 1028–1035.

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 384–397.

Smith, M. A., Roediger, H. L., III, & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1712–1725.

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 80–95.

Stickgold, R. (2013). Parsing the role of sleep in memory processing. *Current Opinion in Neurobiology, 23,* 847–853.

Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience, 3,* 1237–1238.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Talamini, L. M., Nieuwenhuis, I. L., Takashima, A., & Jensen, O. (2008). Sleep directly following learning benefits consolidation of spatial associative memory. *Learning & Memory, 15,* 233–237.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 210–221.

Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: From synaptic and cellular homeostasis to memory consolidation and integration. *Neuron, 81,* 12–34.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie), 56,* 252–257.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language, 50,* 289–335.

Wagner, U., Hallschmid, M., Rasch, B., & Born, J. (2006). Brief sleep after learning keeps emotional memories alive for years. *Biological Psychiatry, 60,* 788–790.

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11,* 571–580.

Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3,* 240–245.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55,* 235–269.

Received September 12, 2017
Revision received December 19, 2017
Accepted January 25, 2018 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

AMERICAN PSYCHOLOGICAL ASSOCIATION

APA **JOURNALS®**

ORDER INFORMATION

Start my 2019 subscription to the
Journal of Experimental Psychology: Learning, Memory, and Cognition® ISSN: 0278-7393

PRICING

APA Member/Affiliate	$212
Individual Nonmember	$562
Institution	$2,091

Call **800-374-2721** or **202-336-5600**
Fax **202-336-5568** | TDD/TTY **202-336-6123**

Subscription orders must be prepaid. Subscriptions are on a calendar year basis. Please allow 4-6 weeks for delivery of the first issue.

Learn more and order online at:
www.apa.org/pubs/journals/xlm

Visit on.apa.org/circ2019
to browse APA's full journal collection.

All APA journal subscriptions include online first journal articles and access to archives. Individuals can receive online access to all of APA's 88 scholarly journals through a subscription to APA PsycNET®, or through an institutional subscription to the PsycARTICLES® database.

XLMA19