Winter 12-19-2019

Accuracy Matters for the Benefits of Sleep After Retrieval Practice

Steven Dessenberger
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Psychological and Brain Sciences

Accuracy Matters for the Benefits of Sleep after Retrieval Practice

by

Steven J. Dessenberger

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

December 2019
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, I would like to thank my advisor, Dr. Mitchell Sommers, for all of his support during this process. I would also like to thank Drs. Henry Roediger and Andrew Butler for serving on my thesis committee. Finally, I would like to thank the National Science Foundation for funding me through their Graduate Research Fellowship Program.

Steve Dessenberger

Washington University in St. Louis

December 2019

Dedicated to Dante, Milo, Lilith, and Anna.

Abstract of the Thesis

Accuracy Matters for the Benefits of Sleep after Retrieval Practice

by

Steve Dessenberger

Master of Arts in Psychology

Washington University in St. Louis, 2019

Professor Mitchell Sommers, Chair

Professor Henry Roediger

Professor Andrew Butler

Previous research suggests that while sleep and retrieval practice can each improve memory on their own, their benefits cannot be combined to produce an additive effect unless feedback is given during the initial test. These previous findings would seem to support a retrieval-as-consolidation of the testing effect, which states that the benefits of retrieval are the result of memory consolidation, a process that normally occurs during the sleep cycle. The present study sought to determine whether the retrieval-as-consolidation account held true when initial test accuracy was considered as a factor. Using foreign language word pairs, we examined the combined effects of sleep, corrective feedback, and initial test accuracy on memory retention after a 12-hour delay. Results suggest that corrective feedback is not necessary to produce a benefit of sleep after a successful retrieval. Feedback only mattered when there was a failure to retrieve during the initial test. These findings suggest that the benefits of retrieval are likely not the result of a memory consolidation process.

Chapter 1: Introduction

A large body of literature has shown that retrieving memory like you would during a test improves memory retention (Adesope, Trevisan, & Sundararajan, 2017; Kang, Gollan, & Pashler, 2013; McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006). Referred to as *retrieval practice*, a typical study demonstrating the benefits of retrieval has participants first learn novel information such as reading a passage of text or studying word pairs. Then participants are either tested on the information (retrieval practice) or perform a control activity such as restudy (studying the information a second time), a distracting activity, or even no activity. Finally, participants are given a final test (often after a delay period). The difference between memory retention for the retrieval practice group and the control group is referred to as the *testing effect* or the benefits of retrieval practice. For example, Glover (1989) demonstrated the benefits of retrieval relative to no activity by having all participants first study essays in preparation for a final test given 4 days later. Half of the participants took a free recall test after 2 days as practice while the remaining participants did nothing. Those that took the practice test remembered more information on the final test despite having the same amount of time to study the provided material.

The mechanisms behind the benefits of retrieval have been the subject of debate, but a recent theory put forward by Antony, Ferreira, Norman, and Wimber (2017) suggests that retrieval practice promotes memory retention because it acts a form of *memory consolidation*. Memory consolidation is a process that is thought to normally occur during the sleep cycle wherein memories are reactivated and then transferred from unstable short-term storage into more stable long-term storage (Diekelmann & Born, 2010). According to Antony et al. (2017), retrieval practice benefits memory retention because like sleep, it consolidates memories.

There are several similarities between the benefits of sleep and retrieval practice that may support a retrieval-as-consolidation account. First, similar to retrieval practice, sleep also has a long history of improving memory retention for studied information (Ellenbogen, Hulbert, Jiang, & Stickgold, 2009; Jenkins & Dallenbach, 1924; van Schalkwijk et al., 2019). For example, Fenn and Hambrick (2013) trained participants on semantically related word pairs to a 60% criterion level. Training either occurred at 9:00 p.m. at night (sleep group) or 9:00 a.m. in the morning (wake group). After a 12-hour delay, participants took a final cued recall test. Researchers found that the sleep group recalled more word pairs compared to the wake group. This means sleep, like retrieval practice, can improve memory retention.

A second similarity between sleep and retrieval practice is both are thought to require memory reactivation to improve memory retention. For retrieval practice, the reactivation of memories is an integral part of the process. During retrieval practice, only items that are retrieved correctly receive a benefit to memory retention while items that are not retrieved correctly receive no benefit and are unlikely to be recalled later (Bjork, 1994; Kornell, Bjork, & Garcia, 2011; Pyc & Rawson, 2009; Smith & Karpicke, 2014). While it is generally accepted that the benefits of retrieval are limited by initial test performance, many studies calculate the benefits of retrieval based on aggregate scores and do differentiate between a retrieval success and a retrieval failure. A meta-analysis of the testing effect conducted by Rowland (2014) found that when initial test performance was less than 50%, the benefits of retrieval were nearly non-existent ($g = .03$). When initial test performance was between 50% and 75%, the benefits of retrieval were considerably larger (g = .29) and only grew in size as initial test accuracy increased. This evidence suggests memories that are not reactivated during retrieval practice do not receive the same benefit to memory retention.

Like retrieval practice, memory consolidation during sleep is also thought to require the reactivation of the memories (Lewis & Durrant, 2011). Direct evidence for this is typically shown in rats wherein neuron firing patterns during learning that are repeated during sleep are associated with increased memory retention (Lee & Wilson, 2002; Wilson & McNaughton, 1994). Unfortunately, tracking memory activation during sleep in humans is not as simple a matter as it is in rats, but it can be simulated if participants were trained using olfactory cues (Rasch, Buchel, Gais, & Born, 2007; Schredl, Hoffmann, Sommer, & Stuck, 2014). For example, in a series of experiments Rasch et al. had participants learn object locations on a screen, which was then followed by a full night's sleep and then a final test of the object locations. In Experiment 1, participants were divided into an odor group and an odorless group. The odor group was presented with the smell of a rose during training and again during sleep. The odorless group were not presented with an odor during training or sleep. At final test, the odor group remembered the object locations better than the odorless group. Researchers concluded that the reintroduction of the odor during sleep cued and reactivated the memory of the location-learning task which facilitated memory consolidation.

Finally, a third similarity between sleep and retrieval practice is that both slow the decay of memories or essentially safeguard against time-based forgetting (Gais, 2006; Talamini, Nieuwenhuis, Takashima, & Jensen, 2008). When memories are created during study, they are thought to be in an unstable state that is prone to decay (forgetting). However, after memory consolidation during sleep, memories are more resistant to the negative effects of decay. Gais, et al. reported that after sleep consolidation, there was no difference in memory retention whether the final test was given 24 or 36 hours later which suggests the memories were being preserved or protected from the negative effects of time-based forgetting. The same kind of protection can be found in the benefits of retrieval practice. For example, Roediger and Karpicke (2006) demonstrated the difference in the rate of memory decay by comparing retrieval practice to restudy after a 5-minute, 2-day, and 1-week retention interval. Although restudy produced the best final test performance for the 5-minute retention interval, retrieval practice produced higher memory retention when measured at the 2-day and 1-week retention intervals because the rate of decay was slower than that of the restudy condition. Essentially, sleep and retrieval practice both slow the rate of memory decay compared to no sleep and restudy, respectively.

These similarities between sleep and retrieval practice have led to speculation that memory retrieval could act as a fast-route to memory consolidation that does not require sleep (Antony et al., 2017). An additional piece of evidence supporting this account can be found when the benefits of sleep and retrieval are combined. If both sleep and retrieval practice consolidate memories, then their combined effects are unlikely to produce an additive benefit compared to either on its own. As noted, consolidation is the process by which memories are transferred from unstable short-term storage into more stable long-term storage. If memories are being consolidated during retrieval practice, then being consolidated again during sleep should not provide a significant benefit to memory retention because the memories are already in stable long-term storage. Consistent with this reasoning, prior research has demonstrated that combining sleep and retrieval practice does not produce better memory retention compared to retrieval practice on its own (Abel et al., 2018; Bäuml, Holterman, & Abel, 2014). Researchers compared 12-hour retention intervals that contained either a period of sleep or uninterrupted wakefulness after a participant performed retrieval practice or restudy (Bäuml et al., 2014). While there was a testing effect present in the wake condition that showed retrieval practice produced better retention compared to restudy, there was no testing effect in the sleep condition. Put another way, sleep improved memory after restudy to the point that the memory retention in the restudy-sleep group was equal to that of both test groups. These findings would seem to support a retrieval-as-consolidation account because while restudied memories benefitted from sleep-consolidation, retrieved memories did not.

However, it is difficult to draw conclusions from Bäuml et al. (2014) alone because a large portion of the research investigating the memory benefits of sleep utilize pretest/posttest designs in their methodology (Ellenbogen et al., 2009; Feld, Weis, & Born, 2016; Fenn & Hambrick, 2013; Payne et al., 2012). If retrieval practice and sleep cannot be combined to produce an additive effect, why then does sleep research that includes retrieval practice prior to sleep routinely demonstrate a benefit of sleep on memory retention? One possible factor is that the pretests in sleep studies nearly always include *corrective feedback* while Bäuml et al. (2014) did not include corrective feedback during retrieval practice.

Corrective feedback is a powerful tool often used during retrieval practice that can increase learning. As previously noted, a failure to retrieve during testing does not provide a benefit to memory retention which means only a subset of the items on the test are likely to be remembered later. One way to get around this limitation of retrieval practice is using corrective feedback during training (Butler & Roediger, 2008; Pashler, Cepeda, Wixted, & Rohrer, 2005). For example, if a test-taker failed to retrieve an answer, feedback in the form of the correct answer could then be given which essentially allows the test-taker to restudy the information and increases the likelihood that they would remember the information later. The potency of corrective feedback was shown by a recent meta-analysis that found retrieval practice followed by feedback produced larger testing effects ($g = .73$) compared to retrieval practice followed by no feedback ($g = .39$; Rowland, 2014).

Abel et al. (2018) demonstrated that the presence of feedback during retrieval practice alters the effects of sleep on memory retention. They compared the effects of sleep to no sleep on restudy, on retrieval practice without feedback and on retrieval practice with feedback. They found that, similarly to Bäuml et al. (2014), sleep improved memory retention for the restudy condition and it did not improve memory retention for the retrieval practice without feedback condition. As for the testing with feedback condition, memory was improved by the inclusion of sleep compared to the no-sleep condition. Abel et al. concluded that while sleep can modulate the benefits of retrieval relative to restudy when feedback was not included, sleep does not modulate the benefits of retrieval relative to restudy when corrective feedback is given during training. To put it simply, sleep produces a benefit after retrieval practice only if feedback was included during training.

Although it is possible that the presence of feedback during training can eliminate sleep's ability to modulate the benefits of retrieval, it is also possible that the interaction demonstrated by Abel et al. (2018) is actually the result of retrieval failure during practice. During testing, it is often the case that some target items are retrieved and some are not. When an item is not retrieved during testing, there is no benefit of retrieval for that item (Kornell et al., 2011; Pyc & Rawson, 2009). Assuming the retrieval-as-consolidation account is true, that would mean for missed items, no consolidation takes place. Items not yet consolidated by retrieval could then potentially still be in unstable short-term storage which would allow sleep to consolidate those memories and transfer them into long-term storage. Therefore, if the retrieval-as-consolidation account is true, previous findings indicating sleep benefits memory after retrieval practice only if feedback is included could potentially be explained by two hypotheses: 1) sleep cannot provide additional benefit to memory after a successful retrieval because consolidated memories cannot benefit from a second-round of consolidation, and 2) sleep does not improve memories for items missed during retrieval when feedback is absent because it cannot consolidate memories that have already been forgotten.

The first hypothesis (consolidated memories cannot be consolidated a second time) is based on the idea that retrieval consolidates the unstable memories created during the initial study period. If ten items are first studied and then only seven of those items are retrieved during retrieval practice, then only those seven items have been consolidated. The remaining three items are still considered unstable and likely to be susceptible to time-based forgetting. Feedback provides an opportunity to restudy the three missed items but as suggested by past research, simple restudy does not provide an opportunity for consolidation (Abel, et al. 2018; Bäuml et al., 2014). This means that memory for those three missed items would be strengthened by feedback but would still remain in unstable storage. This would allow sleep to potentially consolidate those memories. Essentially, only restudied information can be consolidated by sleep, while retrieved memories cannot which would explain why prior research demonstrated large benefits of sleep after restudy and after retrieval practice with feedback but not after retrieval practice when feedback is omitted (Abel et al., 2018; Bäuml et al., 2014).

The second hypothesis (sleep consolidation cannot occur for forgotten items) is based on findings that, in sleep research, the benefits of memory consolidation are tied to the proximity of learning to the onset of sleep (Diekelmann & Born, 2010). People sleep on a daily basis (with some rare exceptions). If sleep consolidated memories from the entire day equally, then it would not matter when information was learned, be it 12-hours before sleep or immediately prior to sleep. However, past research has shown that proximity of learning to sleep matters, as events learned closer in time to sleep are remembered better than those learned longer before sleep (Gais, 2006). One possible explanation for the importance of proximity is that sleep cannot benefit items that have already been forgotten. Items learned earlier in the day undergo more decay due to time prior to the consolidation process compared to items learned later and therefore are more likely already be forgotten by the time sleep occurs. If sleep could consolidate forgotten memories, then not only would proximity not matter, we would also expect retrieval practice without feedback to receive an additional memory benefit from sleep. This is because in retrieval practice paradigms, all items have been studied at least once, meaning a memory has been created for every item on the test. Consider our previous example of ten items learned and seven correctly retrieved during practice. While the seven retrieved items were potentially consolidated by the retrieval process, the remaining three items were not. The three items not retrieved were studied at one point and therefore have a corresponding memory, but that memory had already been forgotten by the time the test began. If sleep could potentially recover forgotten memories, then we would expect to see a benefit of sleep for missed/forgotten items during retrieval practice without feedback, but this is not the case (Abel et al., 2018; Bäuml et al., 2014). Therefore, our second hypothesis is that sleep cannot consolidate memories that have already been forgotten.

Assuming the retrieval-as-consolidation account is true, testing the hypotheses that consolidated memories cannot benefit from consolidation twice and that consolidation cannot occur for forgotten memories requires focused attention on retrieval accuracy during training. A successful retrieval and a failed retrieval are not the same thing as one produces a memory benefit (potentially consolidation) and the other does not. While previous sleep research has opted to use aggregate accuracy at final test or simple difference scores (Final Test Score – Initial Test Score), to test both hypotheses, performance needs to be measured and tracked for each item individually.

In the present study, we compared memory retention for foreign language word pairs after a 12-hour retention interval while looking specifically at the item-level performance during retrieval practice. Participants completed four blocks of study and retrieval practice across three days. During half of the blocks, the retrieval practice was accompanied by immediate feedback while the other half had no feedback. Using a counter-balanced schedule to ensure each type of block was followed by a 12-hour retention interval with one that included sleep and one that did not, we compared final test performance relative to initial test performance measuring at the individual item level. On the assumption that the retrieval-as-consolidation account is true, we predicted that sleep would only provide a memory benefit after retrieval for items that were not correctly retrieved during the initial test when feedback was provided. This means that sleep would not promote retention for items successfully retrieved during retrieval practice regardless of feedback condition nor it would it benefit items missed in the no feedback retrieval practice condition.

Chapter 2: Method

2.1 Participants

A total of 204 participants were recruited from a private U.S. Research University for the study. Five *a priori* established criteria excluded 129 participants from the dataset prior to the primary analysis resulting in a remaining total of 75 participants (51 female, $M_{age} = 20.23$). The target sample size of 75 was determined based on power analysis with effect sizes derived from pilot testing. Of the 5 exclusionary criteria, 4 were taken from Abel, et al. (2018) to include removing participants that reported napping during the day, not sleeping at night, consuming alcohol, or not returning for the test session(s). Due to our use of a within-subjects design, we had a total of 5 sessions and further limited our sample to only participants that completed all 5 sessions at approximately a 12-hour delay as not doing so would introduce the possibility of order effects, fatigue effects, or delay effects that could not be controlled via random assignment. These 5 control measures were implemented in a step-wise fashion (order to follow) resulting in the removal of: (Step 1) 43 participants did not complete all sessions or elected to drop-out prior to completing the experiment (Step 2) 33 participants that did not complete all sessions within the target window of 8-16 hours between each session, (Step 3) 47 participants that slept during one of the day cycles, and (Step 4) 6 participants that consumed alcohol during the course of the experiment. No participants were excluded based on not sleeping during a night-cycle. While this is a high-rate of attrition, online experiments with a large quantity of items have been shown to have larger attrition rates compared to their in-class counterparts and this attrition was comparable to the one we found during pilot testing (Hoerger, 2010).

2.2 Design

A 2x2 within-subjects experimental design investigated the effects of feedback during retrieval (feedback vs no feedback) and sleep between study and test (no sleep vs sleep) on memory retention as measured by a final cued recall test. Half of the participants were randomly assigned to start their first session at 9:00 p.m. and the other half started at 9:00 a.m.

Over the course of 5 sessions spaced 12 hours apart, participants learned 15 Japanese-English word-pairs per session and were tested on those words at the beginning of the next session. This process repeated until participants had learned and were tested on a total of 60 word-pairs (the fifth session only tested the previously learned words and did not add a new set of 15 words). Each learning session introduced the word pairs in a random order one at a time until all 15 had been studied. All learning sessions were then followed by cued recall practice for those 15 words with the Japanese word serving as the cue and English translation serving as the target. Half of the cued recall tests included immediate feedback given after each trial that consisted of the correct Japanese-English word pair appearing on screen for 3s while the other half of the cued recall tests did not include feedback. Learning sessions were counterbalanced for each participant such that one feedback session occurred in the morning and one at night, and one no-feedback session occurred in the morning and one at night.

The final tests were always given 12 hours after a learning session and consisted of presentation of a cue word from the recently learned set of 15 words and request for the target. The order of the words were randomized and feedback was not given during the final test.

2.3 Materials

Sixty Japanese-English word pairs were utilized during this study (for a comprehensive list, see Appendix). Each Japanese word was a two-syllable nouns and none were cognates of their English counterpart. Words were randomly assigned to groups of 15. Counter-balancing ensured that both the order of the presentation of groups of 15 as well as the condition assignment (e.g. sleep and feedback vs no-sleep and no-feedback, etc.) was evenly distributed across participants. Because all participants self-reported as fluent English speakers with no previous experience with Japanese, English is considered a known language (L1) for all participants while Japanese is considered a novel foreign language (L2).

2.4 Procedure

Participants completed a questionnaire that reviewed the schedule of the experiment as well as surveyed the participants prior language experience. Participants were then informed of the general procedure and were made aware that they would be asked to learn a set of words by studying the word pairs followed by a practice test shortly after learning the words (retrieval practice). They were also instructed that they would be given a final test on the words 12 hours later. At this point, half of the participants were assigned to start the initial training at 9:00 a.m. and half were instructed to start at 9:00 p.m.

At the beginning of each session, participants were asked to supply information regarding the previous twelve hours to include any sleep and/or alcohol consumption. During the first of the five total sessions, participants then immediately began the initial training. Regardless of condition, the training process consisted of one round of study and one round of retrieval practice. Participants were informed whether the retrieval practice session would include feedback or not. During the round of study, the participants saw all 15 L2 words paired with their

L1 translation for a period of 10s each in a random order. After studying all 15 words, participants began the single round of retrieval practice. Participants saw the L2 vocabulary words one at a time in a random order. For each word, participants were asked to submit the L1 translation by typing in the answer within 10s. Then, for the feedback conditions, participants would then see the correct L2 and L1 pairing appear on screen for a period of 3s. In the no-feedback condition, participants instead immediately proceeded to the next item.

After a 12-hour delay, participants were tested on the 15 words. For the second session and every session thereafter, this final test took place before the initial training for the next set of words. This was done to reduce the negative effects of interference. During the final tests, participants were given the L2 vocabulary words as a cue and provided with 10s to supply the L1 translation. No feedback was provided during the final test. After this final test, participants were instructed that those 15 words would not be tested again and could be forgotten. This was done to reduce the frequency of intrusions during future testing. Participants then were directed to the initial training for the next set of 15 words and this process repeated until participants had studied, performed retrieval practice, and completed a final test for a total of 60 words.

Chapter 3: Results

3.1 Circadian Effects

To determine if the time of day played a role in learning, a paired t-test compared the initial test scores of tests taken during the morning session to those taken during the night cycle. If a significant difference was found, it would indicate that differences found in final tests may actually be a result of the time of day the information was studied rather than the sleep that occurred during the retention interval. No significant difference was found between initial tests taken in the morning ($M = .47$, $SD = .25$) and those taken at night ($M = .47$, $SD = .27$), $t(149) = .22$ $p = .82$, $d = .03$. Additional figures depicting the descriptive statistics for the initial and final tests can be found in the supplementary figures section.

3.2 Final Test

A hierarchical fixed-slope logistic linear regression model analyzed the influence of sleep (sleep vs no sleep), feedback (feedback vs no feedback) and initial test accuracy (accurate vs inaccurate) on final test performance after a 12-hour delay. The no-sleep condition, the no-feedback condition, and an inaccurate retrieval at initial test were each coded as a 0 in the model while sleep, feedback and an accurate retrieval were each coded as a 1. A table containing the regression coefficients can be found on Table 1 and a visual depiction of the full model can be seen in Figure 1. We included an error term for individual participants in the model as well as for individual words. This allowed the model to account for differences in participant abilities and differences in individual item difficulty. Unless otherwise indicated, reported *p*-values were based on the *z*-distribution and were calculated by extracting coefficients from the model.

Of primary interest, we found a significant two-way interaction between sleep and feedback on final memory retention, ($\beta = .64$, $z = 2.31$, $p = .02$, $OR = 1.90$, 95% CI [1.10, 3.28])[1]. As can be seen in Figure 1, the two-way interaction between feedback and initial test accuracy was also significant ($\beta = -1.63$, $z = 6.39$, $p < .001$, $OR = .20$, 95% CI [.12, .32]) which suggests that while feedback is beneficial to recall after a failed retrieval, it does not improve memory retention when retrieval is accurate. The three-way interaction between sleep, feedback, and initial retrieval accuracy did not reach significance ($\beta = -.60$, $z = 1.66$, $p = .098$, $OR = .55$, 95% CI [.27, 1.12]) nor did the two-way interaction between sleep and initial test accuracy ($\beta = .50$, $z = 1.75$, $p = .08$, $OR = 1.66$, 95% CI [.94, 2.91]). Main effects can be found in Table 1.

To fully investigate our hypotheses as well as these interactions, linear comparisons using a no-intercept version of the model analyzed the simple effects (the following p-values were adjusted using the Holm method; a full list of linear comparisons can be found in the supplementary tables section). We found that after inaccurate initial test retrieval, providing feedback increased the likelihood that the item would be retrieved correctly at final test ($z = 6.45$, $p < .001$, $OR = 3.73$ 95% CI [2.14, 6.51]), but including sleep without feedback did not improve the likelihood of retrieval at final test ($z = .62$, $p = .53$, $OR = 1.15$ 95% CI [.61, 2.18]). However, if both sleep and feedback were provided after an initial retrieval failure, not only was retention improved in comparison to the no-feedback/no-sleep condition ($z = 10.7$, $p < .001$, $OR = 8.22$ 95% CI [4.81, 14.0]), but was also significantly improved the probability of retrieval compared to the no-sleep condition when feedback was included ($z = 5.19$, $p < .001$, $OR = 2.20$ 95% CI [1.45, 3.33]).

---

[1] Odds-ratio (OR) should be interpreted as the odds with which a positive outcome is expected to occur given a one-unit increase in the predictor variable. For example, an OR of 1 is predicting no change in the odds of an accurate answer being retrieved at the final test. As the OR increases beyond 1, the odds of retrieving the correct answer at the final test increases. As the OR decreases to less than 1, the expected odds of retrieving the correct answer decreases.

Essentially, sleep improved memory retention after a failed retrieval attempt only if feedback was provided.

Contrary to our initial hypothesis that consolidated items would not benefit from a second-round of consolidation, for the group of items that were retrieved accurately at initial test, the linear comparisons indicated that sleep did increase the probability of retrieval at final test in comparison to the no sleep condition and this was true for the feedback condition ($z = 4.47$, $p < .001$, *OR* =2.00 95% *CI* [1.31, 3.06]) as well as the no-feedback condition ($z = 3.88$, $p < .001$, *OR* =1.91 95% *CI* [1.21,3.03]). For the no-sleep condition, adding feedback did not significantly improve the probability of retrieval at final test when the initial retrieval was accurate ($z = 2.05$, $p = .08$, *OR* =7.33 95% *CI* [.485, 1.11]).

Chapter 4: Discussion

Consistent with our hypothesis that sleep cannot improve memory retention for already forgotten memories, we found that sleep did not increase the probability of accurate retrieval at final test when the initial retrieval was inaccurate unless feedback was provided. Contrary to our hypothesis that memories consolidated by retrieval could not benefit from the consolidation that takes place during the sleep cycle, we found that sleep did improve memory retention when the initial retrieve was accurate, regardless of feedback condition. Our findings do not fully support a retrieval-as-consolidation account of the testing effect, given the assumption that two forms of consolidation performed on the same memory should not enhance memory retention.

Abel et al. (2018) suggested that sleep's ability to modulate the benefits of retrieval is contingent on the presence of feedback, but as shown by Figure 1, sleep's impact on the benefits of retrieval was the same whether feedback was provided or not. When the initial retrieval was accurate, sleep further enhanced memory retention, regardless of the feedback condition. It is only when initial retrieval was inaccurate that the presence of feedback was crucial to facilitating a benefit from sleep. Since a failure to retrieve does not produce a benefit of retrieval (no testing effect) we would argue that the benefits of sleep and the benefits of retrieval are actually additive regardless of feedback condition.

Our initial hypothesis that consolidated memories would not benefit from a second source of consolidation was our method for testing the retrieval-as-consolidation account. This hypothesis was based on our understanding of the consolidation process which we defined as the transfer of memories from unstable short-term storage into more stable long-term storage. We had assumed that once a memory was in stable long-term storage, sleep could not produce a benefit by consolidating it a second time. While *reconsolidation* of previously consolidated memories is possible, that requires a reminder prior to reconsolidation (for a review, see Klinzing, Rasch, Born, & Diekelmann, 2016; Stickgold & Walker, 2007). The reminder prior to reconsolidation is considered crucial, as it transfers memories in stable long-term storage to a more labile state. Then reconsolidation can transfer the labile memory back into long-term storage, reinforcing memory retention. However, we have no such reminder in this paradigm and even if the immediate feedback condition would constitute a reminder, we observed no difference in the benefits of sleep between the feedback and no feedback groups when initial retrieval was accurate. Since we observed a benefit of sleep after successful retrieval, we would argue that our findings do not fully support a retrieve-as-consolidation account. While there are many similarities in the benefits of retrieval and sleep, we would need to either adjust our definition of sleep consolidation or conclude that the benefits retrieval are derived from a different process.

One limitation of the present study was our use of only one type of stimuli. Foreign language word pairs serve as an excellent medium for memory exercises because they limit the possibility of previous knowledge/associations interfering with learning. However, foreign language stimuli require an additional type of processing in comparison to native language stimuli. According to the type of processing resource allocation (TOPRA) model, native language stimuli require semantic learning while foreign language vocabulary require both form learning and semantic learning (Barcroft, 2002; Sommers & Barcroft, 2013). A foreign language word is comprised of unfamiliar phoneme combinations that need to be integrated with a semantic concept (the native language translation), and this type of learning may alter how it is encoded into long-term storage during retrieval practice. For example, recent research with foreign language word pairs has shown that individual differences in working memory do not modulate the benefits of testing relative to restudy, but research with native language prose material indicates that differences in working memory do modulate the benefits of testing relative to restudy (Agarwal, Finley, Rose, & Roediger, 2017; Bertilsson, Wiklund-Hörnqvist, Stenlund, & Jonsson, 2017; Minear, Coane, Boland, Cooney, & Albat, 2018). Differences in the processing of foreign language vocabulary and native-language material could potentially account for why the present study found a benefit of sleep after retrieval practice when no feedback was given while previous research using native-language word pairs did not (Abel et al., 2018; Bäuml et al., 2014).

A second limitation is our limited understanding of the consolidation process. Because sleep requires a lengthy amount of time to execute, identifying the effects of multiple rounds of consolidation in a relatively short-time period is a challenge. Reconsolidation studies that include multiple periods of sleep typically span multiple days (Klinzing et al., 2016; Stickgold & Walker, 2007). It is possible that multiple rounds of consolidation taking place in a few hours could provide an enhanced benefit and would explain why we found a benefit of sleep and retrieval in the present study. However, to our knowledge, no study has identified the effects of multiple rounds of sleep taking place in the span of a few hours.

A third limitation of the present study was our coding of errors during retrieval practice. A failure to retrieve does not necessarily mean that no information is recalled. When a wrong answer is retrieved during testing, the result is increased memory retention for the incorrectly retrieved answer. This process can create false memories that are likely to be recalled during future testing (Butler, Marsh, Goode, & Roediger, 2006; Fazio & Marsh, 2010; Roediger & Marsh, 2005). During the present study, we coded all answers as either incorrect or correct and did not analyze the effects of a false retrieval. Future research should investigate the influence of false memory creation on the benefits of sleep and retrieval practice.

4.1 Conclusion

Sleep and retrieval practice are both powerful methods for improving memory retention, but they both share the same flaw. Neither can increase memory retention for information that has already been forgotten. In that sense, they are preventative measures for safeguarding memories rather than a method for recovering lost memories. Their benefits can be additive which means taking tests at night prior to sleep may be more beneficial for student learning compared to taking tests in the morning. However, caution must be exercised as a failure to retrieve the correct answer can severely limit learning unless feedback is provided.

References

Abel, M., Haller, V., Köck, H., Pötschke, S., Heib, D., Schabus, M., & Bäuml, K.-H. T. (2018). Sleep reduces the testing effect—But not after corrective feedback and prolonged retention interval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000576

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, *25*(6), 764–771. https://doi.org/10.1080/09658211.2016.1220579

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a Fast Route to Memory Consolidation. *Trends in Cognitive Sciences*, *21*(8), 573–576. https://doi.org/10.1016/j.tics.2017.05.001

Barcroft, J. (2002). Semantic and Structural Elaboration in L2 Lexical Acquisition. *Language Learning*, *52*(2), 323–363. https://doi.org/10.1111/0023-8333.00186

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*(6), 1568–1581. https://doi.org/10.1037/xlm0000025

Bertilsson, F., Wiklund-Hörnqvist, C., Stenlund, T., & Jonsson, B. (2017). The testing effect and its relation to working memory capacity and personality characteristics. *Journal of Cognitive Education and Psychology*, *16*(3), 241–259. https://doi.org/10.1891/1945-8959.16.3.241

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

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, *20*(7), 941–956. https://doi.org/10.1002/acp.1239

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604–616. https://doi.org/10.3758/MC.36.3.604

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*(2), 114–126. https://doi.org/10.1038/nrn2762

Ellenbogen, J. M., Hulbert, J. C., Jiang, Y., & Stickgold, R. (2009). The Sleeping Brain's Influence on Verbal Memory: Boosting Resistance to Interference. *PLoS ONE*, *4*(1), e4117. https://doi.org/10.1371/journal.pone.0004117

Fazio, L. K., & Marsh, E. J. (2010). Correcting False Memories. *Psychological Science*, *21*(6), 801–803. https://doi.org/10.1177/0956797610371341

Feld, G. B., Weis, P. P., & Born, J. (2016). The Limited Capacity of Sleep-Dependent Memory Consolidation. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01368

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review*, *20*(3), 501–506. https://doi.org/10.3758/s13423-012-0366-z

Gais, S. (2006). Sleep after learning aids memory recall. *Learning & Memory*, *13*(3), 259–262. https://doi.org/10.1101/lm.132106

Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*(3), 392–399. https://doi.org/10.1037/0022-0663.81.3.392

Hoerger, M. (2010). Participant Dropout as a Function of Survey Length in Internet-Mediated University Studies: Implications for Study Design and Voluntary Participation in Psychological Research. *Cyberpsychology, Behavior, and Social Networking*, *13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during Sleep and Waking. *The American Journal of Psychology*, *35*(4), 605. https://doi.org/10.2307/1414040

Kang, S. H. K., Gollan, T. H., & Pashler, H. (2013). Don’t just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychonomic Bulletin & Review*, *20*(6), 1259–1265. https://doi.org/10.3758/s13423-013-0450-z

Klinzing, J. G., Rasch, B., Born, J., & Diekelmann, S. (2016). Sleep’s role in the reconsolidation of declarative memories. *Neurobiology of Learning and Memory*, *136*, 166–173. https://doi.org/10.1016/j.nlm.2016.10.004

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*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Lee, A. K., & Wilson, M. A. (2002). Memory of Sequential Experience in the Hippocampus during Slow Wave Sleep. *Neuron*, *36*(6), 1183–1194. https://doi.org/10.1016/S0896-6273(02)01096-6

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, *15*(8), 343–351. https://doi.org/10.1016/j.tics.2011.06.004

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*(2), 200–206. https://doi.org/10.3758/BF03194052

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1474–1486. https://doi.org/10.1037/xlm0000486

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When Does Feedback Facilitate Learning of Words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8. https://doi.org/10.1037/0278-7393.31.1.3

Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for Semantically Related and Unrelated Declarative Information: The Benefit of Sleep, the Cost of Wake. *PLoS ONE*, *7*(3), e33079. https://doi.org/10.1371/journal.pone.0033079

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Rasch, B., Buchel, C., Gais, S., & Born, J. (2007). Odor Cues During Slow-Wave Sleep Prompt Declarative Memory Consolidation. *Science*, *315*(5817), 1426–1429. https://doi.org/10.1126/science.1138581

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Marsh, E. J. (2005). The Positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1155–1159. https://doi.org/10.1037/0278-7393.31.5.1155

Schredl, M., Hoffmann, L., Sommer, J. U., & Stuck, B. A. (2014). Olfactory Stimulation During Sleep Can Reactivate Odor-Associated Images. *Chemosensory Perception*, *7*(3–4), 140–146. https://doi.org/10.1007/s12078-014-9173-4

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, *22*(7), 784–802. https://doi.org/10.1080/09658211.2013.831454

Sommers, M. S., & Barcroft, J. (2013). Effects of Referent Token Variability on L2 Vocabulary Learning: Referent Token Variability and L2 Vocabulary Learning. *Language Learning*, *63*(2), 186–210. https://doi.org/10.1111/lang.12007

Stickgold, R., & Walker, M. P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Medicine*, *8*(4), 331–343. https://doi.org/10.1016/j.sleep.2007.03.011

Talamini, L. M., Nieuwenhuis, I. L. C., Takashima, A., & Jensen, O. (2008). Sleep directly following learning benefits consolidation of spatial associative memory. *Learning & Memory*, *15*(4), 233–237. https://doi.org/10.1101/lm.771608

van Schalkwijk, F. J., Sauter, C., Hoedlmoser, K., Heib, D. P. J., Klösch, G., Moser, D., … Schabus, M. (2019). The effect of daytime napping and full-night sleep on the consolidation of declarative and procedural information. *Journal of Sleep Research*, *28*(1), e12649. https://doi.org/10.1111/jsr.12649

Wilson, M., & McNaughton, B. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*(5172), 676–679. https://doi.org/10.1126/science.8036517

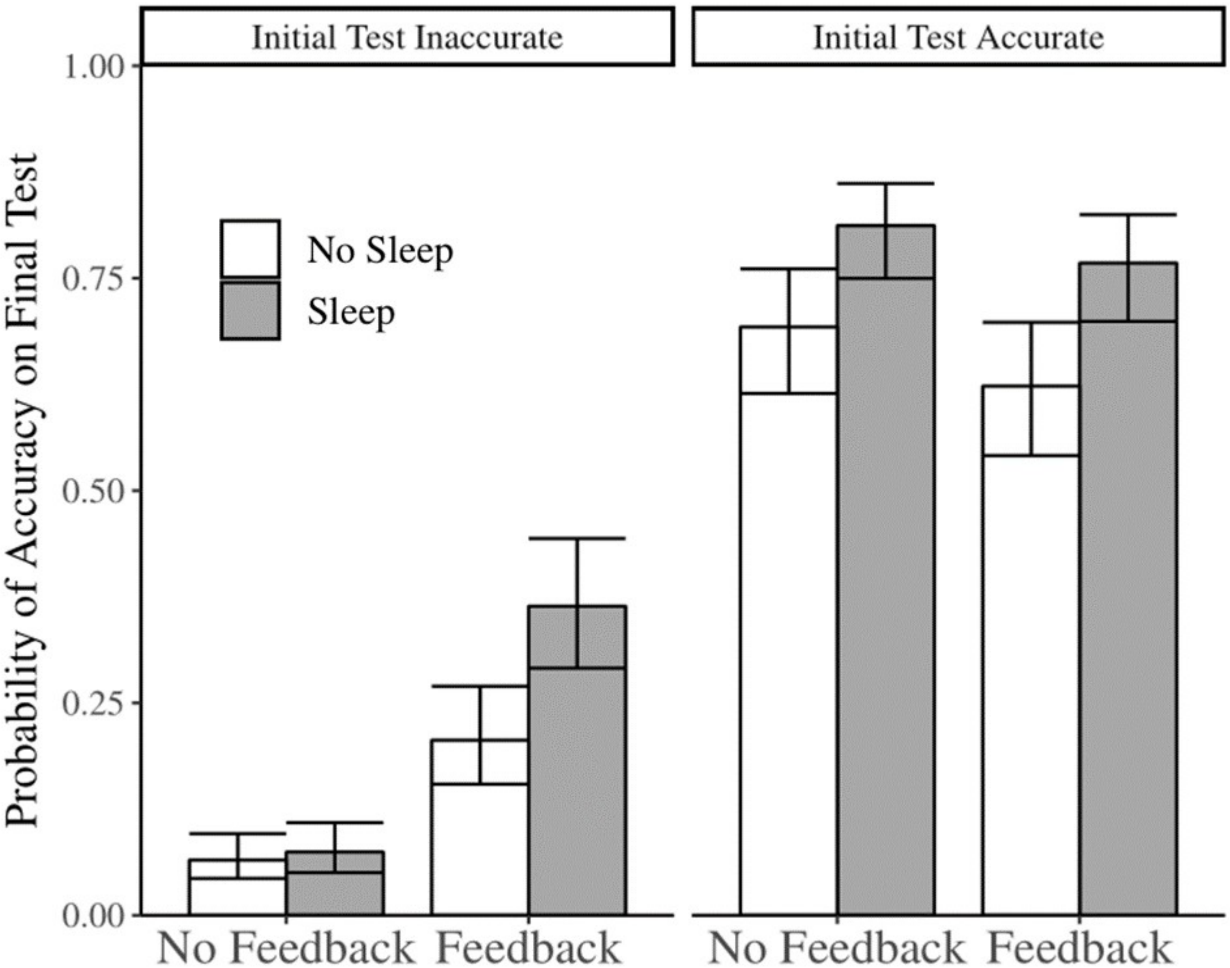
Tables and Figures

Table 1. Logistic Regression Fixed Effects Output measured in Logit Units

	β	Std. Error	*z*-value	*p*-value	*Odds Ratio*	95% CI
Full Model						
Intercept	-2.67	.21	-12.2	< .001*	.06	[.04, .11]
Initial Test Accurate	3.48	.20	17.13	< .001*	32.3	[21.8,48.2]
Sleep	.15	.23	.63	.53	1.15	[.73, 1.82]
Feedback	1.32	.20	6.46	< .001*	3.73	[2.50, 5.57]
Initial Test Accurate x Sleep	.50	.28	1.75	.08	1.66	[.94, 2.91]
Initial Test Accurate x Feedback	-1.63	.25	-6.39	< .001*	.20	[.12, .32]
Sleep x Feedback	.64	.28	2.314	.02*	1.90	[1.10, 3.28]
Initial Test Accurate x Sleep x Feedback	-.60	.36	-1.66	.098	.55	[.27, 1.12]

Notes. Intercept is No Feedback, No Sleep, Initial Test Inaccurate

Figure 1. Predicted probability of final test performance by group



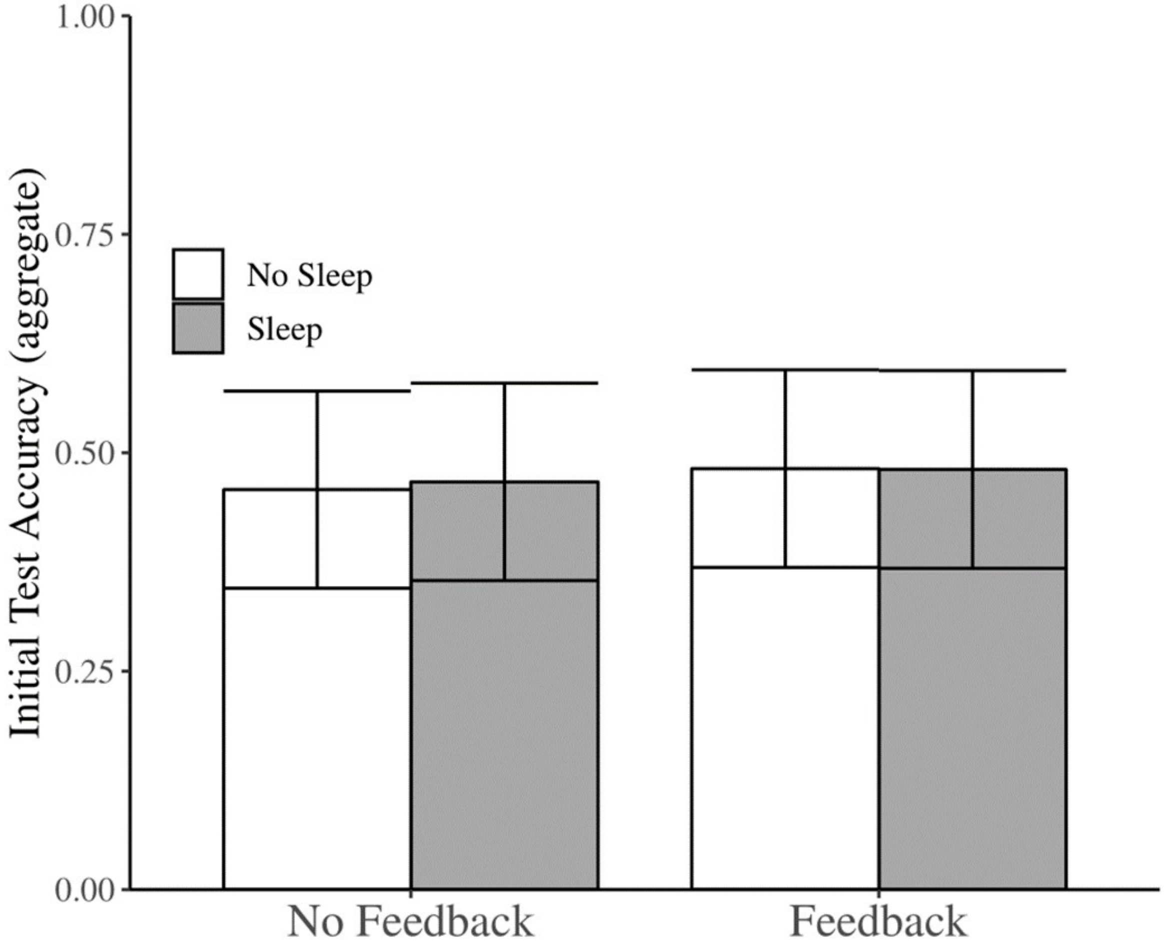
Notes. Error bars are 95% CI

Supplementary Table 1. Linear Comparisons

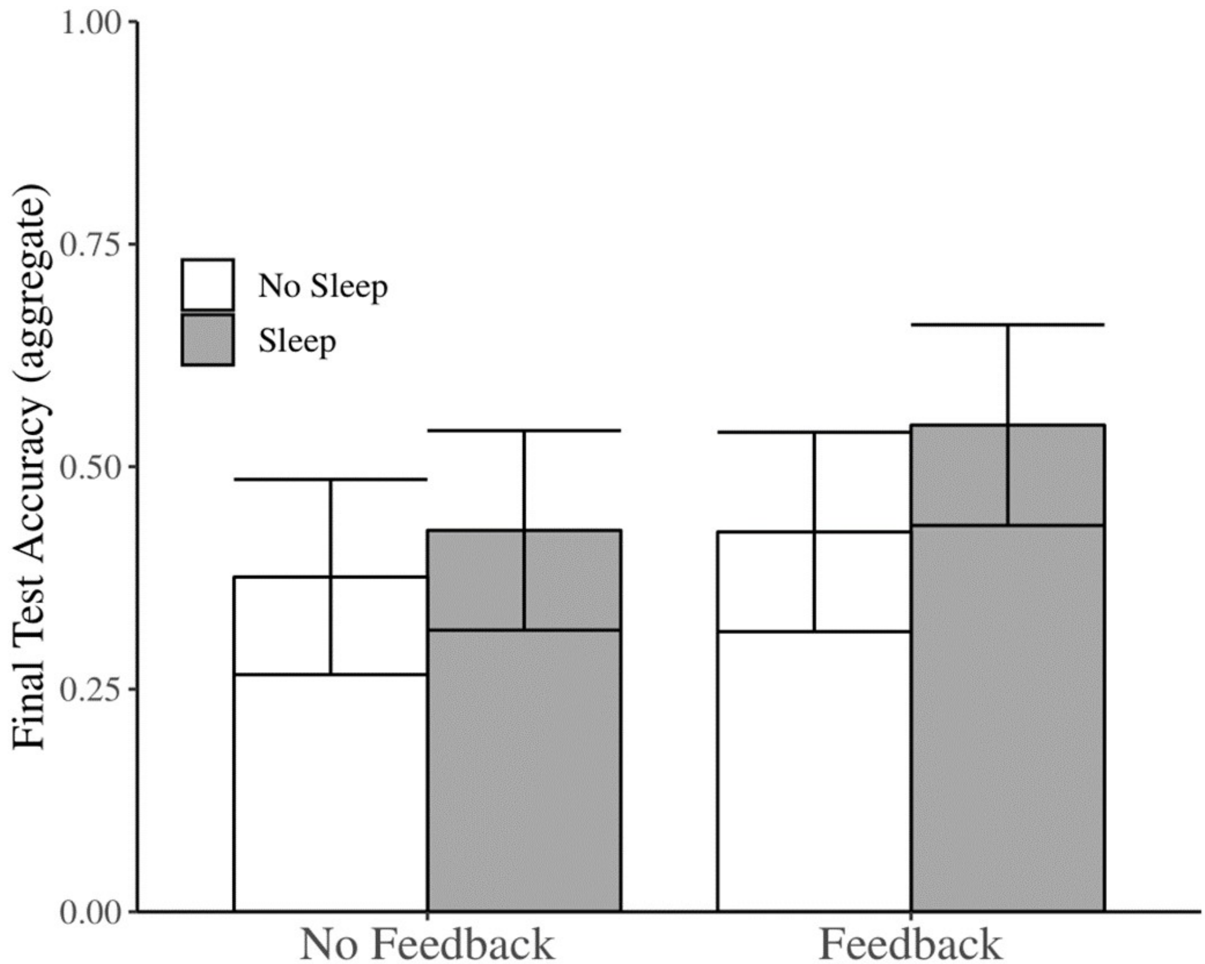
Comparison	*z*-value	*p*-value	*Odds Ratio*	95% CI
Inaccurate: Feedback: No Sleep vs Inaccurate: No Feedback: No Sleep	6.46	<.001*	3.74	[2.14,6.52]
Inaccurate: No Feedback: Sleep vs Inaccurate: No Feedback: No Sleep	.626	.53	1.16	[.61,2.19]
Inaccurate: Feedback: Sleep vs Inaccurate: Feedback: No Sleep	5.19	<.001*	2.20	[1.46,3.33]
Inaccurate: Feedback: Sleep vs Inaccurate: No Feedback: No Sleep	10.7	<.001*	8.23	[4.81,14.1]
Accurate: Feedback: No Sleep vs Accurate: No Feedback: No Sleep	-2.05	.08	.73	[.49,1.11]
Accurate: No Feedback: Sleep vs Accurate: No Feedback: No Sleep	3.88	<.001*	1.92	[1.21,3.03]
Accurate: Feedback: Sleep vs Accurate: Feedback: No Sleep	4.47	<.001*	2.00	[1.31,3.06]
Accurate: Feedback: Sleep vs Accurate: No Feedback: No Sleep	2.38	.052	1.47	[.95,2.28]
Accurate: No Feedback: Sleep vs Inaccurate: No Feedback: Sleep	18.9	<.001*	53.6	[30.2,95.4]

Notes. OR calculated by converting logit differences

Supplementary Figure 1. Average initial test accuracy



Supplementary Figure 2. Average final test performance



Appendix

Japanese	English
Ani	Brother
Batsu	Wrong
Boken	Adventure
Boshi	Hat
Chukan	Middle
Dare	Who
Eiga	Movie
Fure	Boat
Gane	Money
Gohan	Rice
Heiho	Square
Heiwa	Peace
Hinkon	Poverty
Hito	Person
Iho	Illegal
Ijo	Over
Inu	Dog
Jihi	Mercy
Jiro	Castle
Jisho	Dictionary

Japanese	English
Jiyu	Freedom
Kaiyo	Ocean
Kami	Paper
Konran	Confusion
Konton	Chaos
Kuma	Bear
Kutsu	Shoes
Mado	Window
Neko	Cat
Nihon	Japan
Niku	Meat
Nishi	West
Nodo	Throat
Otto	Husband
Ringo	Apples
Saino	Talent
Sake	Salmon
Seicho	Growth
Seiko	Success
Seki	Cough

Japanese	English
Senbo	Envy
Shika	Dentist
Shinko	Belief
Shinpai	Worry
Shippa	Failure
Shita	Under
Shori	Victory
Shukan	Habit
Shukyo	Religion
Suimin	Sleep
Tama	Ball
Tango	Word
Tomi	Wealth
Tsuma	Wife
Uma	Horse
Ushi	Cow
Yoku	Bath
Yubi	Finger
Yuki	Courage
Zubon	Pants