pii: sp-00579-15 http://dx.doi.org/10.5665/sleep.6092

SLEEP COGNITION AND BEHAVIOR

Sleep Restriction Impairs Vocabulary Learning when Adolescents Cram for Exams: The Need for Sleep Study

Sha Huang, PhD[1]; Aadya Deshpande, BS[1]; Sing-Chen Yeo, BS[1]; June C. Lo, PhD[1]; Michael W.L. Chee, MBBS[1]; Joshua J. Gooley, PhD[1,2]

[1]*Center for Cognitive Neuroscience, Program in Neuroscience and Behavioral Disorders, Duke-NUS Medical School, Singapore;* [2]*Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore*

Study Objectives: The ability to recall facts is improved when learning takes place at spaced intervals, or when sleep follows shortly after learning. However, many students cram for exams and trade sleep for other activities. The aim of this study was to examine the interaction of study spacing and time in bed (TIB) for sleep on vocabulary learning in adolescents.

Methods: In the Need for Sleep Study, which used a parallel-group design, 56 adolescents aged 15–19 years were randomly assigned to a week of either 5 h or 9 h of TIB for sleep each night as part of a 14-day protocol conducted at a boarding school. During the sleep manipulation period, participants studied 40 Graduate Record Examination (GRE)-type English words using digital flashcards. Word pairs were presented over 4 consecutive days (spaced items), or all at once during single study sessions (massed items), with total study time kept constant across conditions. Recall performance was examined 0 h, 24 h, and 120 h after all items were studied.

Results: For all retention intervals examined, recall of massed items was impaired by a greater amount in adolescents exposed to sleep restriction. In contrast, cued recall performance on spaced items was similar between sleep groups.

Conclusions: Spaced learning conferred strong protection against the effects of sleep restriction on recall performance, whereas students who had insufficient sleep were more likely to forget items studied over short time intervals. These findings in adolescents demonstrate the importance of combining good study habits and good sleep habits to optimize learning outcomes.

Keywords: sleep, spaced learning, vocabulary learning, declarative memory, adolescents

Citation: Huang S, Deshpande A, Yeo SC, Lo JC, Chee MW, Gooley JJ. Sleep restriction impairs vocabulary learning when adolescents cram for exams: the Need for Sleep Study. *SLEEP* 2016;39(9):1681–1690.

Significance

Many adolescents are exposed to sleep restriction during school days. This behavior could have negative consequences for learning of skills and facts, hence eroding the value of formal education. Most research on the role of sleep in declarative memory has focused on recall performance after a single study opportunity, while in real life students usually learn across days. We found that adolescents exposed to sleep restriction exhibited greater forgetting of items that were learned during single study sessions, but not when items were studied over multiple days. Students with insufficient sleep who cram for exams might be especially prone to forgetting newly learned material, but deficits in learning can be minimized by spacing study sessions over time.

INTRODUCTION

Many students are habitually exposed to sleep restriction during school days.[1–4] In the United States, the majority of adolescents aged 15–17 years report sleeping 7 h or less per night,[5] while the National Sleep Foundation recommends 8–10 h of sleep for this age group.[6] The high prevalence of sleep restriction is alarming because deficits in learning and related cognitive faculties such as attention occur following insufficient sleep.[7,8] Additionally, when sleep is sacrificed for other late-night activities, test scores and performance on assignments on the following day generally suffer.[9] Furthermore, many students procrastinate on studying for exams,[10] which can lead to cramming at the expense of sleep. Presumably, a more effective strategy for long-term learning would be to space studying over multiple days with intervening episodes of sufficient sleep.

More than 130 years ago, Hermann Ebbinghaus conducted a series of pioneering studies on learning and memory in which he described the basic properties of learning and forgetting curves.[11] As part of this work, Ebbinghaus showed that long-term retention is improved when studying takes place at spaced intervals, as opposed to all at once. Hence, simply re-distributing study time can improve memory, even when total study time is kept constant. The advantages of increased study spacing for long-term learning (i.e., the spacing effect) have been demonstrated for a wide variety of learning tasks ranging from fine motor skills to vocabulary learning, assessed either in the laboratory or in the classroom.[12–15] In his analyses of retention as a function of time, Ebbinghaus found that recall performance showed little drop-off between 9 h to 24 h after learning (i.e., spanning the night), as compared to retention assessed before and after this interval. His findings were later replicated and extended by others,[16,17] demonstrating that retention is better if followed by sleep rather than wakefulness.[7,18] For decades, it was assumed that less forgetting occurs across a night of sleep because there is less memory interference compared to the daytime.[19] However, recent work suggests that memories are actively strengthened and stabilized during sleep.[20]

Despite the benefits of increased study spacing and sleep on memory, little is known about how these factors interact to influence long-term learning. In a meta-analysis of spaced learning on verbal memory tasks, it was shown that for retention intervals ranging from a day to a month, the spacing effect was greatest when the interval between study sessions was about a day.[12] Because the focus of these studies was on the effects of study spacing, the potential role of sleep in learning was not considered. Recently, it was shown that long-term retention of Swahili-English translations was better when sleep occurred between evening and morning study sessions, as compared to an equivalent period of wakefulness between morning and evening study sessions.[21] These results indicate that sleep between encoding and restudying improves long-term memory. However, it has yet to be examined whether repeated exposure to sleep restriction, which often occurs during the school week in adolescents, modulates the effects of study spacing on memory.

The goal of the present study was to characterize the interaction between sleep duration and study spacing on long-term vocabulary learning. We hypothesized that if sleep accounts for a large portion of the spacing effect, then individuals exposed to sleep restriction would show a smaller benefit of increased study spacing on long-term learning, as compared to individuals with longer sleep. Alternatively, if exposure to partial sleep deprivation leads to greater forgetting of items that are studied only once rather than spaced across days, then increasing the degree of study spacing would result in a larger benefit on vocabulary learning in individuals exposed to sleep restriction.

METHODS

Subjects and Recruitment

Healthy adolescents (n = 60) aged 15–19 years took part in the Need for Sleep Study[22] during their vacation period (November 26 to December 9, 2014). Most students were recruited during open-invitation information sessions held at their respective schools, while others responded to advertisements or learned about the study through their peers. Interested students attended a study briefing session with at least one of their parents. Informed written consent was obtained from at least one parent with assent provided by his/her child. Study procedures were approved by the National University of Singapore Institutional Review Board, and research was compliant with ethical principles outlined in the Declaration of Helsinki.

Potential subjects completed a series of questionnaires to determine their eligibility and health status. The Beck Anxiety Inventory was used to exclude individuals with moderate to severe anxiety levels (score ≥ 16),[23] and the Beck Depression Inventory II was used to exclude participants with moderate to severe depressive symptoms (score ≥ 20).[24] Subjects were ineligible if they had a history of psychiatric illness, central nervous system disease, or organ disease (e.g., renal or liver impairment). Participants underwent a face-to-face interview to ascertain that they would be comfortable living with relative strangers in a community setting for a 2-week period. Additional exclusionary criteria included high risk for sleep apnea assessed using the Berlin Questionnaire,[25] a body mass index ≥ 30 kg/m^2, travel across more than 2 time zones in the month before the study, or consumption of ≥ 5 caffeinated beverages per day. Subjects also completed several questionnaires that were not used to determine eligibility, including the Pittsburgh Sleep Quality Index to assess sleep quality in the past month[26]; the Horne-Östberg Morningness-Eveningness Questionnaire to determine chronotype[27]; the Epworth Sleepiness Scale to assess excessive daytime sleepiness[28]; the Chronic Sleep Reduction Questionnaire for measuring symptoms of chronic sleep restriction[29]; and Raven's Advanced Progressive Matrices for evaluating non-verbal intelligence.[30] Individuals who completed the initial screening visit were required to wear an actigraphy device (Actiwatch 2, Philips Respironics Mini-Mitter, Bend, OR) on their non-dominant wrist for 1 week during their school term. Participants were excluded if they were habitual short sleepers, defined as having an average daily time in bed (TIB) < 6 h per night with less than 1 h of sleep extension on weekends versus weekdays.

Protocol Overview

Subjects took part in a 14-day study conducted at a boarding school in Singapore. In the week before the study, subjects were required to maintain a fixed sleep-wake schedule with 9 h of TIB per night from 23:00 to 08:00, which was verified by actigraphy monitoring. To accommodate study planning, participants were randomly assigned to 9-h TIB and 5-h TIB groups (n = 30 per group) prior to the start of the 14-day protocol; however, they were not informed of their assignment until the first day of the study. Of the 30 subjects who were assigned to the 9-h TIB group, 2 subjects withdrew before the start of the 14-day study, and 1 participant withdrew during the second baseline day due to personal reasons. Another subject in the 9-h TIB group completed the study but was excluded post hoc when it was determined that he did not comply with the sleep schedule. Therefore, the 9-h TIB group had 26 participants (11 males, mean ± standard deviation [SD] = 16.8 ± 1.2 years) and the 5-h TIB group had 30 participants (14 males, mean ± SD = 16.4 ± 0.9 years).

The 14-day protocol consisted of 3 nights of baseline sleep, 7 nights of sleep manipulation, and 3 nights of recovery sleep (Figure 1). Subjects arrived at the boarding school in the morning and were oriented to study procedures on the first day of the protocol. During the baseline nights, all subjects were given a 9-h opportunity for sleep from 23:00 to 08:00. Over the next 7 nights, students who were assigned to the 9-h TIB group kept the same sleep-wake schedule (Figure 1A), whereas the 5-h TIB group was allowed to sleep from 01:00 to 06:00 each night (Figure 1B). On the final 3 nights of the study, all participants were given a 9-h sleep opportunity from 23:00 to 08:00. Subjects were discharged from the study on the following day.

Study Environment

The study took place at the boarding school while school was out of session, and hence the facilities described here were used primarily by study participants. Subjects remained on site for the full 2 weeks, and were under constant supervision by the researchers. During their free time, students spent most of their time in a large common room that had both natural and electric lighting. They were permitted to play games, read, review school work, and use their personal electronic devices for entertainment. Subjects were also allowed to interact freely with other study participants and research staff members. Napping, caffeinated beverages, and strenuous physical activity were prohibited. Subjects were served breakfast, lunch, and dinner in a school cafeteria, and had free access to snacks during their free time in the common room.

On each day of the study, participants completed a computer-based neurobehavioral test battery (~30 min) in the morning, afternoon, and evening (10:00, 15:00, and 20:00), as described in our previous work.[22] Several stand-alone tests were also administered during the course of the study, including the vocabulary learning task described here. All testing took place in a standard classroom setting with assigned seating. A study laptop was assigned to each student, which allowed all subjects to complete tests at the same time.

At night, students slept in separate beds in a residence building with 2 students per room. Each room was air-conditioned and had an attached bathroom. Male and female students were housed in different residential blocks, and the different sleep groups (9 h TIB and 5 h TIB) were housed on different floors. The windows in each bedroom were covered by the investigators to minimize the amount of sunlight in the morning. Subjects were also provided with ear plugs to reduce noise and were allowed to adjust the temperature of their rooms. Polysomnographic data were analyzed on selected nights of the study (nights 3, 4, 7, 10, 11, and 13) to validate the sleep manipulation. As reported elsewhere, both sleep groups slept a similar amount at baseline, whereas the 5-h TIB condition was associated with a large daily reduction in total sleep time relative to the 9-h TIB condition.[22,31] Bedtimes and wake times were enforced by researchers who stayed in the residence halls, and students were asked to surrender their personal electronic devices during sleep periods.

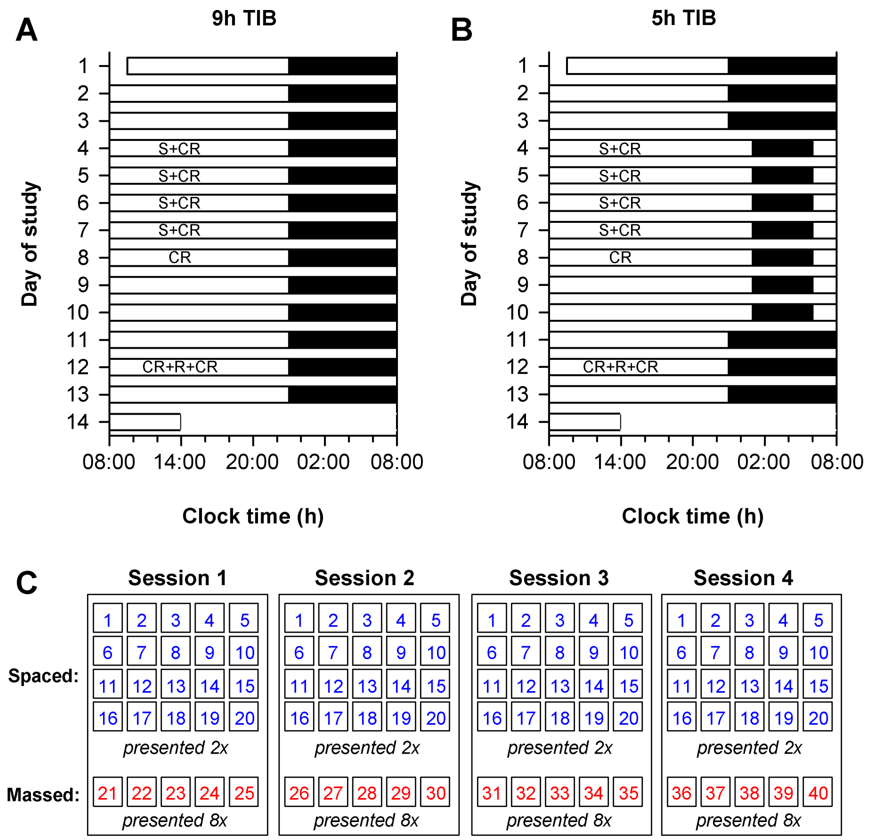

Figure 1—Sleep schedule and flashcard spacing. Adolescents took part in a 14-day research study carried out at a boarding school. After 3 baseline nights with 9 h of time in bed (TIB) for sleep, participants were exposed to 7 nights with either **(A)** 9 h of TIB from 23:00 to 08:00, or **(B)** 5 h of TIB from 01:00 to 06:00, followed by 3 recovery nights with 9 h of TIB for sleep. Over a period of 4 consecutive days (days 4–7), students completed a vocabulary learning task in which they studied 40 Graduate Record Examination (GRE)-type words. Cued recall tests were given immediately after each study session, and also 24 h and 120 h after studying (days 8 and 12). After the test given on day 12, students were given a one-time review session that was followed by another cued recall test. In panels A and B, S = study sessions; CR = cued recall tests; and R = review. **(C)** For each participant, flashcards were randomly partitioned into a stack of 20 cards shown twice per day across each of the study sessions (spaced items), and 4 smaller stacks with a different set of cards shown each day (massed items). Spaced items and massed items were presented in different blocks, with the order of presentation randomized and counterbalanced in each sleep group. Panel C is modified from Kornell, Applied Cognitive Psychology, 2009.[32]

Vocabulary Learning Task and Study Spacing

Subjects were oriented to the vocabulary learning task on the fourth day of the study (i.e., after the last baseline night), with instructions given to all participants at the same time. Studying and testing occurred in the same classroom with assigned seating, and laptops were used to administer the task. Participants were made fully aware of the study and testing schedule at the start of the experiment. The vocabulary learning task took place at 13:45 each day over 4 consecutive days (days 4 to 7). Across the 4 study sessions, subjects studied 40 word pairs using digital flashcards (Figure 1C). Each word pair consisted of a Graduate Record Examination (GRE)-type English word (the cue word) followed by a synonym (the target word). We chose to use GRE-type words, rather than unrelated word pairs, to ensure better ecological validity. The flashcard learning paradigm was based on a previous study that used the same set of word pairs, which were selected because they represent the types of words given on standardized tests such as the GRE.[32]

Subjects were instructed that they would be tested later on a cued recall test, during which they would have to report the matching word for each of the 40 cue words. During study sessions, the "front" of each flashcard (the cue word) was shown for 5 seconds (e.g., encomium:______), followed by the "back" of the flashcard (the target word) for another 5 seconds (e.g., ______:praise). This was followed immediately by presentation of the next word pair. Digital flashcards were presented using E-Prime 2.0 Professional software (Psychology Software Tools, Inc, Sharpsburg, PA).

Study spacing was manipulated by using flashcard stacks of different sizes. All subjects were exposed to spaced and massed study conditions, with the order of presentation randomly assigned and counterbalanced within each sleep group (9 h TIB and 5 h TIB). In each subject, the 40 word pairs were randomly partitioned into one stack of 20 flashcards (spaced items) and 4 smaller stacks of 5 flashcards each (massed items) (Figure 1C). Once the order of words within a flashcard stack was assigned, it remained fixed for a given participant. Spaced items were presented twice during each of the 4 study sessions, whereas massed items were studied 8 times during a single session, with a new stack of 5 flashcards presented each day. Therefore, subjects studied 25 flashcards per day, comprising 20 spaced items and 5 massed items (Figure 1C). By the end of the fourth study session, all word pairs were presented 8 times, with 80 seconds of total viewing time for each flashcard (40 seconds each for the front and back).

Cued Recall Test

During cued recall tests, participants were shown all 40 cue words in random order. Each cue word was displayed for a fixed period of 15 seconds, during which subjects were instructed to type the target word beneath the cue word using their laptop keyboard. Tests were given immediately after each of the study sessions to track vocabulary learning (days 4–7). The rationale for testing all 40 GRE-type cue words, even when subjects had yet to be exposed to every flashcard, was to evaluate whether participants were able to guess the target words based on their pre-existing knowledge. In addition to assessing recall immediately after each study session, cued recall tests were given 24 h and 120 h after the final study session (days 8 and 12). Directly after the test that occurred at the 120-h retention interval, subjects were given a one-time review session in which all 40 flashcards were shown once in random order. Similar to study sessions, the front and back of each flashcard were shown for 5 seconds each. The review session was followed by a final cued recall test to evaluate the benefit of the review session on recall performance.

Analyses and Statistics

For each of the cued recall tests, responses were analyzed using an algorithm that counted both correctly spelled answers and misspelled correct answers as correct responses (Lenient scorer program, N. Kornell).[33] The score on each test was converted to a percentage (correct responses/number of tested items × 100%) prior to performing statistical analyses. Test scores were analyzed separately for the 3 different phases of the experiment: study, retention, and review. Performance during the study phase was evaluated using cued recall tests that were given immediately after each of the 4 study sessions (days 4–7). Retention was evaluated after all 40 flashcards had been studied, but without additional review, corresponding to tests that occurred immediately after the fourth study session (0 h, day 7), on the following day (24 h, day 8), and 5 days after the last study session (120 h, day 12). Finally, the benefit of reviewing the flashcards was evaluated by comparing test scores before and after the one-time review session that occurred 120 h after the last study session.

During the study phase, statistical comparisons were not performed for spacing effects because the number of spaced items and massed items differed during the first 3 study sessions, and were equal only after the fourth study session. We therefore provide only descriptive results for the effects of study spacing during the study phase of the experiment, e.g., general characteristics of the learning curves. For data collected during the retention and review phases of the study, statistical comparisons were performed using ANOVA with 4 factors including TIB (9 h versus 5 h), flashcard spacing (spaced versus massed), order of spacing (spaced first versus massed first), and session/day of testing. TIB and order of spacing were between-subjects factors, and test session and flashcard spacing were within-subjects factors. We anticipated that these factors might interact in complex ways to affect cued recall performance. Because we did not have a strong theoretical basis for including some interactions and dropping others from the model, all possible interaction terms were considered in the ANOVA. Analyses were performed using the generalized linear model function command in R version 3.1.1,[34] with a Gaussian distribution and identity link function. Model parameters were determined by maximum likelihood estimation, and deviance was computed to measure how closely the predicted values from the fitted model matched the actual data. The significance of each main/interaction effect was assessed by comparing the deviance for the full model with the deviance for the reduced model (i.e., the nested model) with that particular effect removed. Because the sampling distribution of deviance can be approximated by a χ^2 distribution, statistical comparisons between models were performed using a likelihood ratio χ^2 test ($LR\ \chi^2$). Multiple comparison testing was performed using a Tukey Honestly Significant Difference test at a family-wise confidence level of 95%, and effect size was measured using eta-squared (η^2) (R version 3.1.1).

RESULTS

Subjects who were assigned to the 9-h and 5-h TIB groups were similar for all measures taken during the screening process, including basic subject characteristics, sleep behavior, anxiety and depression scores, and nonverbal intelligence (Table 1). Based on actigraphy-estimated sleep duration during the school term, subjects assigned to both sleep groups slept about 2 h more on weekends versus school days. After a 1-week washout period during the school holiday period with a prescribed 9-h sleep schedule (23:00 to 08:00), students participated in the 14-day study at the boarding school. Effects of sleep duration and study spacing (spaced versus massed) on learning of GRE-type words were examined by exposing adolescents to a week of sleep restriction (5 h TIB) or an age-appropriate amount of sleep (9 h TIB) after each night of studying (Figure 1).

First, we examined learning curves based on cued recall performance on all 40 items tested *immediately* after each of the 4 study sessions that occurred on consecutive days. Both sleep groups showed a nonlinear increase in recall performance for spaced items across the 4 study sessions (Figure 2A), with the largest improvement observed between the first and second study sessions. There was an approximately linear increase in test scores for massed items because participants were shown 5 new massed word pairs each day but were tested on all items from the beginning of the experiment. Because subjects were tested on some massed items on days 4–6 (study sessions 1–3) that had not been encoded yet, we performed a secondary analysis which only involved test scores on those massed items

Table 1—Subject characteristics.

Measure	9h TIB (n = 26) Mean	SD	5h TIB (n = 30) Mean	SD	t/χ^2	P
Age (years)	16.81	1.17	16.43	0.94	1.33	0.19
Sex (% males)	42.30	–	46.70	–	0.11	0.74
BMI (kg/m^2)	20.38	2.55	20.43	2.88	0.07	0.94
BAI score	6.58	4.83	7.80	6.45	0.79	0.43
BDI score	5.19	4.68	6.90	5.49	1.24	0.22
RAPM score	10.38	1.06	9.77	1.98	1.43	0.16
ESS score	6.19	3.57	7.77	3.59	1.64	0.11
MEQ score	49.96	7.15	47.90	7.43	1.05	0.30
CSRQ score	33.81	5.13	34.50	5.77	0.47	0.64
PSQI score	4.58	2.58	5.17	2.32	0.90	0.37
Weekday TST (h)	5.37	0.73	5.61	0.86	1.11	0.27
Weekend TST (h)	7.53	1.14	7.46	1.10	0.21	0.84

TIB, time in bed for sleep; BMI, body mass index; BAI, Beck Anxiety Inventory; BDI, Beck Depression Inventory II; RAPM, Raven's Advanced Progressive Matrices; ESS, Epworth Sleepiness Scale; MEQ, Horne-Östberg Morningness-Eveningness Questionnaire; CSRQ, Chronic Sleep Reduction Questionnaire; PSQI, Pittsburgh Sleep Quality Index; TST, total sleep time, estimated using actigraphy devices when school was in session; SD, standard deviation.

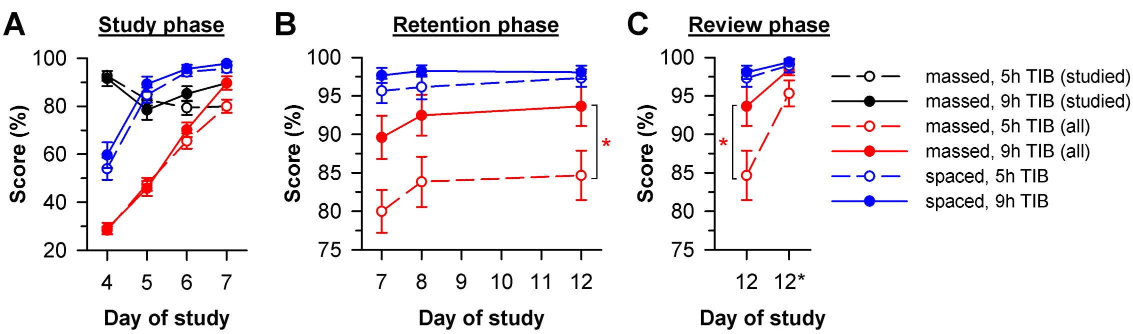

Figure 2—Vocabulary learning and cued recall performance in response to sleep restriction. **(A)** Learning curves for recall of spaced items and massed items (20 items per condition) are shown for groups of adolescents exposed to either 9 h of time in bed (TIB) or 5 h of TIB for sleep after each day of learning. Students were tested immediately after each study session. For items that were massed, test scores are shown separately for those items that had already been studied (black traces), and for all 20 cue words (red traces) including those items that had yet to be encoded. **(B)** After spaced items and massed items were studied an equal number of times, there was a significant interaction between study spacing and TIB ($LR\,\chi^2 = 8.40$, $P < 0.01$), whereby sleep restriction was associated with a significant decrease in cued recall performance on massed items (indicated by the red asterisk, $P < 0.05$), but not on spaced items. **(C)** Following a review session on day 12 in which students were shown all word pairs one more time, a final cued recall test was administered (day 12*) in which test scores on massed items improved substantially. The mean ± SEM is shown for performance on each cued recall test. Additional statistical comparisons during the retention phase and review phase are summarized in Table 2 and in the main text.

that had been studied prior to testing (Figure 2A). By the end of the second study session, test performance was qualitatively similar on spaced items versus massed items, even though subjects had to remember twice as many spaced items with only half the number of flashcard presentations (20 spaced items shown 4 times each versus 10 massed items shown 8 times each). As expected, test scores for massed items and spaced items (20 per condition) were most similar by the end of the fourth study session, after all items had been studied an equal number of times.

Next, we evaluated cued recall performance across different retention intervals (0 h, 24 h, and 120 h). Despite being exposed to an additional 4 days of either 9 h of TIB or 5 h of TIB after the final study session, neither sleep group showed evidence of forgetting (Figure 2B). There was a significant interaction between study spacing and TIB for sleep (Table 2), such that test performance on massed items was poorer in individuals who underwent sleep restriction (Tukey test, $P < 0.001$), while test performance on spaced items was similar between sleep groups (Tukey test, $P = 0.80$; Figure 2B). Consequently, individuals exposed to sleep restriction exhibited a larger improvement in test scores with increased study spacing (Figure 3A). The magnitude of the spacing effect, defined as the difference in test scores for each participant on spaced items versus massed items, was about 2-fold greater in individuals exposed to sleep restriction (Figure 3B). Although there was a main effect of order on test scores with better recall of items presented in the second study block of each study session, there

Table 2—Statistical summary of factors affecting cued recall performance.

Factor	Retention Phase			Review Phase		
	$LR\ \chi^2$	P	η^2	$LR\ \chi^2$	P	η^2
Spacing	61.65	< 0.001*	0.150	21.28	< 0.001*	0.077
TIB	17.29	< 0.001*	0.041	7.55	< 0.01*	0.027
Session	3.19	0.20	0.008	14.67	< 0.001*	0.053
Order	14.96	< 0.001*	0.035	8.75	< 0.01*	0.031
Spacing × TIB	8.40	< 0.01*	0.020	5.06	0.025*	0.018
Spacing × Session	1.30	0.52	0.003	6.85	< 0.01*	0.025
Spacing × Order	3.02	0.083	0.007	3.37	0.066	0.012
TIB × Session	0.092	0.96	< 0.001	1.53	0.22	0.006
TIB × Order	0.064	0.80	< 0.001	0.066	0.80	< 0.001
Session × Order	0.065	0.97	< 0.001	0.60	0.44	0.002

As part of a 14-day protocol, adolescents studied Graduate Record Examination (GRE)-type words over 4 consecutive days, and cued recall performance was tested across different retention intervals (0 h, 24 h, and 120 h; days 7, 8, and 12 of the protocol), and before and after a vocabulary review (120 h; day 12 of the protocol). For each phase of the experiment, a generalized linear model (GLM) was used to examine effects of study spacing (spaced versus massed), time in bed (TIB) for sleep (9 h versus 5 h), test session, and order of study blocks (massed items first versus spaced items first) on cued recall performance. Spacing and Session were within-subjects factors, and TIB and Order were between-subjects factors. All possible interactions were included in the GLM. For each effect, the likelihood ratio χ^2 ($LR\ \chi^2$) statistic was used to evaluate whether the goodness-of-fit for the reduced model (i.e., with that effect removed) differed significantly from the full model with all predictor variables included. Results are shown for main effects and 2-way interactions. All 3-way interactions and the 4-way interaction did not reach statistical significance (Retention phase: $LR\ \chi^2 \leq 0.20$, P > 0.8; Review phase: $LR\ \chi^2 < 1.3$, P > 0.2 for all 3-way/4-way interactions). Asterisks (*) indicate significant effects (P < 0.05), and effect sizes were estimated by eta-squared (η^2).

was no significant interaction involving order and other predictors (Table 2). Likewise, all 3-way interactions and the 4-way interaction did not reach statistical significance ($LR\ \chi^2 \leq 0.20$, P > 0.8 for all 3-way/4-way interactions).

After a one-time review session in which subjects were given an opportunity to study all 40 flashcards one more time, there remained a significant interaction between study spacing and TIB in which recall performance for massed items was worse in individuals exposed to 5 h of TIB versus 9 h of TIB (Tukey test, P < 0.01), while test scores were similar between sleep groups for spaced items (Tukey test, P = 0.99). Additionally, there was a significant interaction between study spacing and test session (Table 2), whereby recall performance on massed items improved after the review session, reaching a similar level as for spaced items (Tukey test, P = 0.51) (Figure 2C). All 3-way and 4-way interaction effects were not significant ($LR\ \chi^2 < 1.3$, P > 0.2 for all interactions).

Because cued recall performance was different between sleep groups only for the massed study condition, we conducted an exploratory analysis to assess whether testing performance differed for massed items presented during each of the different study sessions (Figure 4). First, we evaluated whether overnight forgetting differed between sleep groups during the study phase for different sets of massed items. Based on an ANOVA with the factors Set (massed set 1, 2, 3, and 4), Session (immediate versus 24 h later after a night of sleep), and TIB (9 h versus 5 h), there was a significant interaction between Set and TIB ($LR\ \chi^2 = 8.48$, P = 0.037, $\eta^2 = 0.017$), such that the difference in recall performance between sleep groups was marginally greater for massed items presented in Set 3 (Tukey test, P = 0.024) relative to Sets 1, 2, and 4. Session did not interact with TIB ($\chi^2 = 1.70$, P = 0.19), however, and no other interactions reached statistical significance. There was a main effect of Session ($LR\ \chi^2 = 41.1$, P < 0.0001, $\eta^2 = 0.081$) in which participants showed a drop in test scores from immediate recall to 24 h after the respective massed items were studied (e.g. from day 4 to day 5 for set 1, and from day 5 to day 6 for set 2; Figure 4). Next, we conducted a separate ANOVA to examine recall performance for different sets of massed items during the retention phase of the experiment, i.e., 0 h, 24 h, and 120 h after all flashcard items were studied. There was a significant interaction between Set and TIB ($LR\ \chi^2 = 10.71$, P = 0.013, $\eta^2 = 0.014$); recall of the first set of massed items was similar for the 9h TIB and 5h TIB groups (Tukey test, P > 0.99), whereas participants who underwent sleep restriction had lower test scores for massed items that were presented on the second and third study sessions (Set 2: Tukey test, P < 0.01 and Set 3: P < 0.001). Although the difference in test scores between sleep groups did not reach statistical significance for massed items presented during the fourth study session (Set 4: Tukey test, P = 0.15), the overall time-course of recall performance was comparable to massed items presented on the second and third study sessions (Figure 4).

DISCUSSION

Our results show that daily exposure to partial sleep deprivation impaired vocabulary learning when studied word pairs were massed rather than spaced across consecutive days. Spaced learning conferred strong resistance to the effects of sleep restriction on memory, such that cued recall performance was similar in students exposed to 9 h of TIB versus 5 h of TIB. Our results demonstrate that students exposed to sleep restriction can minimize the negative effects of sleep insufficiency on vocabulary learning by increasing their study spacing.

Sleep Duration Modulates Effects of Study Spacing on Learning

Prior work has shown that memories can be strengthened by increasing the spacing between study sessions, or by allotting oneself adequate sleep. Our data extend these findings by showing that study spacing and sleep duration interact to influence vocabulary recall. The negative impact of sleep restriction on cued recall performance was greater for items that were presented in the massed condition. This could be related to decrements in attention and effort during encoding and retrieval. In the same group of students, we found that exposure to sleep restriction resulted in a greater number of attentional failures on a visual reaction time test.[22] Because students were presented with a greater number of repetitions for massed items during each study session compared with spaced items, attention and effort to study massed items might have diminished by a greater amount in sleep-restricted individuals once these items became highly familiar,[35,36] especially if students thought that they had already mastered the material.[37] In addition, exposure to 5 h of TIB might have impaired overnight consolidation of massed items, as there was a trend for decreased recall performance on the day after encoding relative to students given 9 h of TIB (Figure 4).

Almost all prior research on sleep-dependent learning and consolidation of verbal memory has examined recall performance after a single study session, even though this is not how students learn ordinarily.[38] In contrast to our results for massed items that were studied this way, we found that sleep restriction had no measurable effect on cued recall performance for items that were spaced across 4 consecutive days, despite a 28-h reduction in the opportunity for sleep in the 5-h TIB condition from the first day of studying to the final day of testing. The protective effect of spaced studying during sleep loss could be explained by the benefit of increased retrieval difficulty on declarative memory. When items are studied only once (i.e., massed), exposure to insufficient sleep would be expected to reduce encoding efficiency and increase retrieval difficulty, resulting in poorer recall. However, when studying is spaced out over time, increased retrieval difficulty at encoding may lead to strengthening of memory traces when these items are restudied, thus facilitating subsequent recall.[39–42] This mechanism could serve to protect against the otherwise impairing effects of sleep loss on learning and memory consolidation. It is also possible that the protective effect of increased study spacing is explained in part by increased contextual variability at the time of encoding when items are spaced rather than massed, leading to more cues available to aid in recall.[43]

Interestingly, memory of massed items shown on the first day of studying was similar between sleep groups, whereas sleep restriction was associated with greater forgetting of massed items presented during other study sessions (Figure 4). Because both sleep groups were given 9 h of TIB on the night before the first study session, the strength of encoding would be expected to be similar for the first set of massed items. By comparison, sleep loss on subsequent nights might have resulted in weaker encoding of short-term memory,[44] even if participants were able to perform well on the immediate recall test, thus increasing the likelihood of forgetting of these items. Alternatively, it is possible that 2 nights of sleep restriction represented

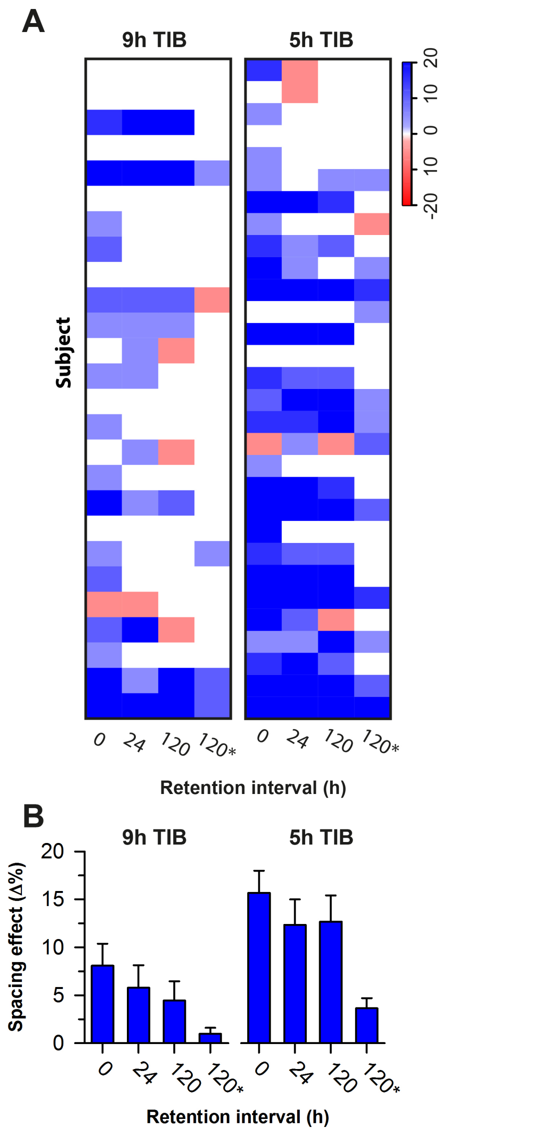

Figure 3—Effects of study spacing on vocabulary learning were greater during sleep restriction. Cued recall performance for Graduate Record Examination (GRE)-type words was examined across different retention intervals for students who were given either 9 h of time in bed (TIB) for sleep each night, or 5 h of TIB. **(A)** The spacing effect, defined as the difference in test scores for spaced items versus massed items, is shown for each student across different retention intervals. In the heat map, blue indicates better cued recall performance on spaced items, and red indicates higher test scores on massed items. Cued recall performance was assessed immediately after the last study session, and also 24 h and 120 h after all items were shown. This was followed by a review session in which students were shown all word pairs once, after which students were given a final cued recall test (120*h). **(B)** The magnitude of the spacing effect was greater in students given 5 h of TIB versus 9 h of TIB (Spacing × TIB: $LR\ \chi^2 = 8.40$, $P < 0.01$; See Table 2). The mean ± SEM is shown. The individual data in Panel A are aligned with group-averaged data in Panel B.

a critical threshold affecting sleep-dependent memory consolidation, whereas students were able to tolerate the first night of partial sleep loss without detectable memory impairment.[45] Although not tested in the present study, sleep might have

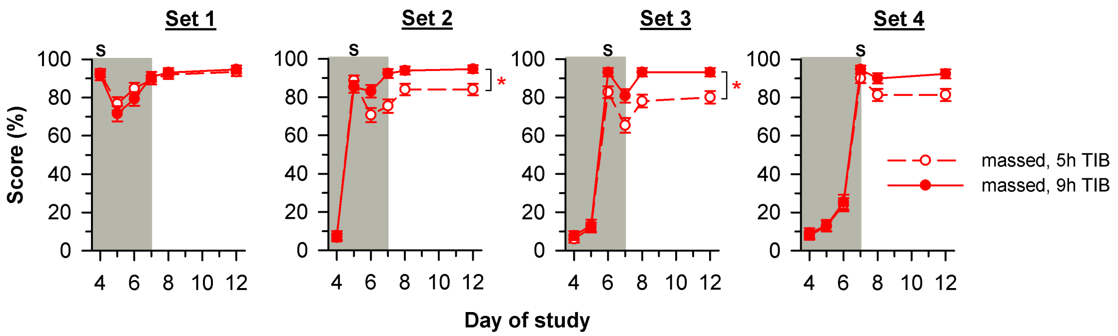

Figure 4—Sleep restriction impaired recall of massed items. Students who were exposed to either 9 h of time in bed (TIB) or 5 h of TIB for sleep used digital flashcards to study Graduate Record Examination (GRE)-type words on 4 consecutive days (i.e., the study phase of the experiment, indicated by the gray boxes). In the massed study condition, a different set of 5 word pairs was presented on each day (Set 1 – day 4; Set 2 – day 5; Set 3 – day 6; Set 4 – day 7) and studied only once, with 8 repetitions per flashcard. For the first set of massed items that was studied, the time course of cued recall performance was similar between sleep groups. In contrast, test scores for massed items shown during the other study sessions were consistently worse in adolescents exposed to sleep restriction. There was a significant interaction between Set and TIB ($LR\ \chi^2 = 10.71$, $P = 0.013$), in which students who were exposed to sleep restriction showed significantly lower test scores for massed items presented in Sets 2 and 3 after all items had been studied (indicated by red asterisks, $P < 0.05$). Additional statistical comparisons are presented in the main text. For each set of massed items, "S" indicates recall performance immediately after each set of items was studied. The mean ± SEM is shown for performance on each cued recall test.

contributed differentially to the initial consolidation and later refinement of memory as participants were exposed to spaced items over the course of 4 consecutive days. For example, it has been shown that as a person masters a complex procedural task that requires training across multiple days (the Tower of Hanoi task), changes in EEG-derived sleep features associate with improvements from initial training to mastery of the skill.[46] At present, however, it remains unclear whether sleep contributes to day-to-day improvements in declarative memory consolidation beyond the first night after encoding.

Based on prior work on forgetting curves, we expected that our participants would show a decrease in cued recall performance as the lag between studying and testing was increased. Contrary to our expectations, after the last study session there was no evidence of forgetting in either sleep group for retention intervals up to 5 days after learning. The difference in cued recall performance between the 5-h TIB and 9-h TIB groups persisted until the one-time review session that occurred 120 h after learning, which then boosted recall for massed items. Our finding that exposure to 5 h of TIB did not result in greater forgetting after all items had been studied is consistent with prior work in adolescents demonstrating that recall of word pairs learned prior to sleep restriction was intact.[47] However, it is possible that the lack of forgetting in our study could be attributed to students learning the word pairs too well for a decrease in cued recall performance to be observed in the time window examined. This possibility could be addressed in future studies by increasing the difficulty of the task, e.g., by increasing the number of word pairs, or by reducing the number of study sessions or number of times that the flashcards are shown.

Study Limitations and Considerations

The present study was not designed to tease apart the differential effects of sleep restriction on encoding versus memory consolidation. To examine this, it would be necessary to compare results for experiments in which encoding is either preceded by sleep restriction and followed by normal sleep, or preceded by normal sleep and followed by exposure to sleep restriction. By comparison, our subjects were exposed to partial sleep deprivation both before and after encoding of vocabulary items. Additionally, the number of massed word pairs learned and the amount of sleep deprivation increased across study sessions. Although this study design simulates what often occurs in adolescents during a typical school week, in which sleep deprivation accumulates over multiple days, we cannot determine with certainty which specific memory processes were affected by exposure to sleep restriction.

We found that the spacing effect was nearly 2-fold greater in the group of students exposed to sleep restriction, but the magnitude of the spacing effect likely depends on numerous factors, including the number and difficulty of items studied, the number of repetitions for each item, the degree of spacing between items during a given study session (determined by the size of the flashcard stack), the time duration between study sessions, and the length of the retention interval. Since only one combination of these factors was examined, the spacing effect might differ if any one of these parameters is manipulated. The limited benefit of increased study spacing in the group of participants given 9 h of TIB could also be explained in part by a ceiling effect, since these students performed well at recalling items that were massed and had less room for improvement relative to the group that underwent sleep restriction. Additionally, recall performance on spaced items was very high for both sleep groups, suggesting that these items might have been over-studied. This might have prevented us from detecting deficits in recall for spaced items in the group of adolescents exposed to 5 h of TIB that would otherwise be observed if the task was made more difficult. Another limitation of our study is that we did not perform an a priori power analysis to assess whether our sample size would be adequate for the statistical comparisons that were performed. It is therefore possible that our sample size was insufficient for detecting differences in test scores for some of the variables and interactions that were examined.

To examine the effects of study spacing on a per individual basis, all students were exposed to massed items and spaced items in separate blocks. Consistent with prior work, an order effect was observed (Table 2), such that word pairs were better learned when presented in the second study block.[32] In our study design, the order of massed items and spaced items was counterbalanced, and effects of order were taken into account in the statistical analyses. The order of presentation did not interact with study spacing (Table 2), indicating that our findings for TIB and study spacing were not confounded by order effects.

In our study, participants were shown all 40 cue words during each cued recall test, including the study phase of the experiment during which subjects were given 5 new massed items to study per day. Therefore, participants were tested on items after each of the first 3 study sessions that had yet to be encoded. Although this allowed us to examine whether students were able to guess target words before they were studied, we do not know whether advanced exposure to unpaired cue words affected subsequent learning and recall of these massed items. Upon close inspection of cued recall performance for massed items (Figure 4), it can be seen that test scores improved across study sessions by a small but consistent amount, even for items that had yet to be studied. After an initial drop in recall performance on the day after each set of massed items was studied, there was a similar increase in test scores on subsequent days when memory was tested. A possible explanation for the improvement in test scores is that some students discussed the study material with each other or looked up tested items using their personal electronic devices during their free time, even though they were explicitly instructed not to do this. Since we aimed to create a set of environmental conditions similar to what students experience during their school term, with studying and testing in a classroom setting, sleep in residence halls, and free time to interact with each other, it was neither practical nor desirable for us to isolate students or ban their use of personal laptops or smartphones. Importantly, the small increase in test scores for non-studied items was similar between sleep groups, and hence did not explain the interaction between study spacing and TIB for sleep.

Implications of Study Spacing and Sleep Restriction for Students

A major goal of education is to promote long-term learning and retention of skills and facts. The spacing effect has obvious implications for students and instructors alike, as learning and retention can potentially be optimized by appropriately spacing study sessions and course material over time.[15,48] Almost all students use some type of spaced learning, but the spacing between studied items, either within or between sessions, can differ markedly. We examined flashcard learning because this is a common approach used by students to prepare for standardized tests including the GRE, and yet there is little guidance for students on how to use flashcards for learning.[49] Large stacks of cards increase spacing between repeated items, whereas smaller decks reduce spacing. Interestingly, most students choose to spontaneously cut large flashcard decks into smaller stacks,[50] even though such behavior likely represents a suboptimal strategy for long-term learning.[32] Moreover, when asked to predict their recall performance of studied items that were massed or spaced, students often misjudge cramming as the better study strategy, even when their test results show the opposite effect.[32]

As suggested in the present study, vocabulary learning is especially impaired when poor study strategies are combined with insufficient sleep. There are many factors contributing to chronic sleep loss in adolescents including early school start times, caffeine consumption in the evening, late-night electronic media use, and school workload.[51] These factors often lead to a disparity between sleep behavior on weekdays versus weekends, and hence social and biological time, which has been dubbed "social jet lag."[3] Similar to the effects of rapid travel across time zones, social jet jag and sleep debt can result in fatigue and suboptimal performance. This potentially leads to poorer academic outcomes,[52] with negative implications for long-term learning. In students who habitually sleep less on school days versus free days, we show that exposure to sleep restriction reduces flashcard learning of GRE-type words when studying is not spaced out over time. These results indicate that cramming for exams and trading sleep for other activities may come at the expense of increased forgetting, whereas increasing study spacing can serve to minimize the negative effects of sleep loss on learning.

REFERENCES

1. Dewald JF, Meijer AM, Oort FJ, Kerkhof GA, Bögels SM. The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: a meta-analytic review. Sleep Med Rev 2010;14:179–89.
2. Gradisar M, Gardner G, Dohnt H. Recent worldwide sleep patterns and problems during adolescence: a review and meta-analysis of age, region, and sleep. Sleep Med 2011;12:110–8.
3. Wittmann M, Dinich J, Merrow M, Roenneberg T. Social jetlag: misalignment of biological and social time. Chronobiol Int 2006;23:497–509.
4. Wolfson AR, Carskadon MA. Sleep schedules and daytime functioning in adolescents. Child Dev 1998;69:875–87.
5. Sleep in America Poll; Sleep in the Modern Family. Arlington, VA: National Sleep Foundation, 2014.
6. Hirshkowitz M, Whiton K, Albert SM, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 2015;1:40–3.
7. Diekelmann S, Born J. The memory function of sleep. Nat Rev Neurosci 2010;11:114–26.
8. Lim J, Dinges DF. A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. Psychol Bull 2010;136:375–89.
9. Gillen-O'Neel C, Huynh VW, Fuligni AJ. To study or to sleep? The academic costs of extra studying at the expense of sleep. Child Dev 2013;84:133–42.
10. Solomen LJ, Rothblum ED. Academic procrastination: frequency and cognitive-behavioral correlates. J Counsel Psychol 1984;31:503–9.
11. Ebbinghaus H. Memory; a contribution to experimental psychology. Teachers College, Columbia University, 1885.
12. Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: a review and quantitative synthesis. Psychol Bull 2006;132:354–80.

13. Donovan JJ, Radosevich DJ. A meta-analytic review of the distribution of practice effect: now you see it, now you don't. J Appl Psychol 1999;84:795–805.
14. Lee TD, Genovese ED. Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. Res Q Exerc Sport 1988;59:277–87.
15. Seabrook R, Brown GD, Solity JE. Distributed and massed practice: from laboratory to classroom. Appl Cognit Psychol 2005;19:107–22.
16. Heine R. Uber Wiedererkennen und ruckwirdende Hemmung. Zeitschrift fur Psychologie und Physiologie der Sinnesorgane 1914;68:161–236.
17. Jenkins JG, Dallenbach KM. Obliviscence during sleep and waking. Am J Psychol 1924:605–12.
18. Gais S, Lucas B, Born J. Sleep after learning aids memory recall. Learn Mem 2006;13:259–62.
19. Wixted JT. The psychology and neuroscience of forgetting. Annu Rev Psychol 2004;55:235–69.
20. Ellenbogen JM, Payne JD, Stickgold R. The role of sleep in declarative memory consolidation: passive, permissive, active or none? Curr Opin Neurobiol 2006;16:716–22.
21. Bell MC, Kawadri N, Simone PM, Wiseheart M. Long-term memory, sleep, and the spacing effect. Memory 2014;22:276–83.
22. Lo JC, Ong JL, Leong RL, Gooley JJ, Chee MW. Cognitive performance, sleepiness, and mood in partially sleep deprived adolescents: the Need for Sleep Study. Sleep 2016;39:687–98.
23. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56:893–7.
24. Beck AT, Steer RA, Brown GK. Beck depression inventory-II. San Antonio, TX: Psychological Corporation, 1996.
25. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999;131:485–91.
26. Buysse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28:193–213.
27. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 1975;4:97–110.
28. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14:540–5.
29. Meijer AM. Chronic sleep reduction, functioning at school and school achievement in preadolescents. J Sleep Res 2008;17:395–405.
30. Raven JC, Court JH. Raven's progressive matrices and vocabulary scales. Oxford Psychologists Press, 1998.
31. Ong JL, Lo JC, Gooley JJ, Chee MW. EEG changes across multiple nights of sleep restriction and recovery in adolescents: the Need for Sleep Study. Sleep 2016;39:1233–40.
32. Kornell N. Optimising learning using flashcards: spacing is more effective than cramming. Appl Cognit Psychol 2009;23:1297–317.
33. Kornell N. Lenient Scorer. 2009. http://sites.williams.edu/nk2/software/.
34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
35. Dellarosa D, Bourne LE Jr. Surface form and the spacing effect. Mem Cognit 1985;13:529–37.
36. Greene R. Spacing effects in memory: evidence for a two-process account. J Exp Psychol Learn Mem Cogn 1989;15:371–7.
37. Zechmeister E, Shaughnessy J. When you know that you know and when you think that you know but you don't. Bull Psychon Soc 1980;15:41–4.
38. Ribiero S, Stickgold R. Sleep and school education. Trends Neurosci Educ 2014;3:18–23.
39. Bjork RA, Allen TW. The spacing effect: consolidation or differential encoding? J Verbal Learning Verbal Behav 1970;9:567–72.
40. Cuddy LJ, Jacoby LL. When forgetting helps memory: an analysis of repetition effects. J Verbal Learning Verbal Behav 1982;21:451–67.
41. Bjork RA, Bjork EL. A new theory of disuse and an old theory of stimulus fluctuation. From learning processes to cognitive processes: Essays in honor of William K. Estes 1992;2:35–67.
42. Gardiner FM, Craik FI, Bleasdale FA. Retrieval difficulty and subsequent recall. Mem Cognit 1973;1:213–6.
43. Lohnas LJ, Polyn SM, Kahana MJ. Contextual variability in free recall. J Mem Lang 2011;64:249–55.
44. Yoo SS, Hu PT, Gujar N, Jolesz FA, Walker MP. A deficit in the ability to form new human memories without sleep. Nat Neurosci 2007;10:385–92.
45. Kopasz M, Loessl B, Valerius G, et al. No persisting effect of partial sleep curtailment on cognitive performance and declarative memory recall in adolescents. J Sleep Res 2010;19:71–9.
46. Fogel SM, Ray LB, Binnie L, Owen AM. How to become an expert: a new perspective on the role of sleep in the mastery of procedural skills. Neurobiol Learn Mem 2015;125:236–48.
47. Voderholzer U, Piosczyk H, Holz J, et al. Sleep restriction over several days does not affect long-term recall of declarative and procedural memories in adolescents. Sleep Med 2011;12:170–8.
48. Moulton CA, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect? A randomized, controlled trial. Ann Surg 2006;244:400–9.
49. Wissman KT, Rawson KA, Pyc MA. How and when do students use flashcards? Memory 2012;20:568–79.
50. Salisbury DF, Klein JD. A comparison of a microcomputer progressive state drill and flashcards for learning paired associates. J Computer-Based Instruction 1988;15:136–43.
51. Wolfson AR, Carskadon MA. Understanding adolescent's sleep patterns and school performance: a critical appraisal. Sleep Med Rev 2003;7:491–506.
52. Touitou Y. Adolescent sleep misalignment: a chronic jet lag and a matter of public health. J Physiol Paris 2013;107:323–6.

ACKNOWLEDGMENTS

The authors thank research assistants, students, and post-doctoral fellows in the Center for Cognitive Neuroscience who assisted in carrying out the study, including subject recruitment, technical support, and supervision of research participants. We thank Dr. James Cousins for providing valuable comments on an earlier draft of the manuscript.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication October, 2015
Submitted in final revised form April, 2016
Accepted for publication May, 2016

Address correspondence to: Joshua J. Gooley, PhD, Duke-NUS Medical School, 8 College Road Singapore 169857; Tel: 65 6516 7430; Fax: 65 6221 8625; Email: joshua.gooley@duke-nus.edu.sg

DISCLOSURE STATEMENT

This was not an industry supported study. This work was supported by the National Medical Research Council, Singapore (NMRC/STaR/0004/2008 and NMRC/STaR/015/2013); the Duke-NUS Signature Research Program funded by the Agency for Science, Technology and Research, Singapore, and the Ministry of Health, Singapore; and the Far East Organization. The work was performed at the Duke-NUS Medical School, Singapore. The authors have indicated no financial conflicts of interest.