Does subjective sleep quality predict cognitive performance? – Evidence from three empirical studies

Zsófia Zavecz[1,2], Adrienn Galkó[3], Dezso Nemeth[2,3]*, Karolina Janacsek[2,3]

[1]Doctoral School of Psychology, Eötvös Loránd University

[2] Brain, Memory and Language Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

[3]Department of Clinical Psychology and Addiction, Institute of Psychology, Eötvös Loránd University, Budapest, Hungary

*Corresponding author:

Email: nemeth@nemethlab.com (DN)

Abstract

The role of sleep in cognitive performance has gained increasing attention in neuroscience and sleep research in the recent decades, however, the relationship between subjective sleep quality and cognitive performance has not yet been comprehensively characterized. In this paper, our aim was to test – across three studies – the relationship between subjective sleep quality and a wide range of cognitive functions in healthy young adults. Sleep quality was assessed by Pittsburgh Sleep Quality Index, Athens Insomnia Scale, and a sleep diary to capture general subjective sleep quality, and Groningen Sleep Quality Scale and a questionnaire about the night prior to testing to capture prior night's sleep quality. Within cognitive functions, we tested several sub-processes of procedural learning, short-term memory, working memory and executive functions (assessed by the Alternating Serial Reaction Time, Digit Span, Counting Span tasks, and the Wisconsin Card Sorting Test, respectively). To test the *amount of evidence* either for associations or no associations between subject sleep quality and cognitive performance, we calculated Bayes Factors along with frequentist analyses. Finally, to further increase the reliability of our conclusions, we also conducted a mini-meta-analysis on the three individual studies included in our paper. Our results, unequivocally across all analyses, show that there is no association between subjective sleep quality and cognitive performance in the domain of procedural learning, short-term/working memory, and executive functions. Our paper can contribute to a deeper understanding of subjective sleep quality and its measures, and we discuss various factors which may affect whether or not associations can be observed between subjective sleep quality and cognitive performance.

Keywords: sleep, cognitive functioning, sequence learning, Alternating Serial Reaction Time (ASRT), Pittsburgh Sleep Quality Index (PSQI)

Introduction

The role of sleep in cognitive performance has been gain increasing attention in neuroscience and sleep research in the recent decades [1,2]. Numerous experimental methods exist that can be employed for examining the association between sleep and cognitive performance. For instance, sleep parameters can be evaluated based on actigraph or electroencephalograph measurements, which are time-consuming and require hardly accessible equipment. Hence, many researchers turn to questionnaires to assess sleep parameters and their relationship with cognitive performance [3-6].

One of the most widely-used sleep questionnaires is the Pittsburgh Sleep Quality Index [PSQI, 7], a self-administered questionnaire, in which participants rate their subjective sleep quality based on several questions, including the average amount of sleep during night, the difficulty falling asleep, and other sleeping disturbances. Nevertheless, there are other popular measurements, such as Athens Insomnia Scale [AIS, 8], measuring difficulties in falling asleep or maintaining sleep and sleep diaries, which accurately capture the sleeping habits of the participants spanning a few days or weeks. Sleep questionnaires and sleep diaries are two different types of self-report measures: while a sleep questionnaire is retrospective, administered at a single point in time, and asks about various aspects of the sleep experience “in general,”, sleep diary is an ongoing, daily self-monitoring. Libman et al. [9] showed that even though results of questionnaires and diaries are highly correlating, there are differences in the means of the sleep parameters depending on the type of self-report measurement, suggesting that the two measurement modalities were tapping the same domains, but yet they capture somewhat different aspects of subjective sleep quality.

Previous research on subjective sleep quality and cognitive performance has led to mixed findings. While some studies have shown that poorer sleep quality measured by the PSQI score was associated with weaker complex working memory performance [4], executive functions [3] and decision-making [5], others have failed to find association between subjective sleep quality and cognitive performance [6,10]. To the best of our knowledge, no studies have focused on the associations between other subjective sleep questionnaires and cognitive performance.

In previous investigations focusing on the association between subjective sleep quality and various aspects of cognitive performance, the potential relationship with procedural learning/memory has largely been neglected. The procedural memory system underlies the learning, storage, and use of cognitive and perceptual-motor skills and habits [11]. Evidence suggests that the system is multifaceted in that it supports numerous functions that are performed automatically, including sequences, probabilistic categorization, and grammar, and perhaps aspects of social skills [12-16]. In light of the importance of this memory system, the clarification of its relationship with sleep would be indispensable. Importantly, we are not aware of any studies focusing on the relationship between subjective sleep quality and procedural learning.

In this paper, our aim was to – across three studies – provide an extensive investigation on the relationship between subjective sleep quality and performance in various cognitive functions, including procedural learning in healthy young adults. Subjective sleep quality was assessed not only by PSQI but also by other, less frequently used sleep quality measures: namely, Athens Insomnia Scale (AIS, Study 1-3), Groningen Sleep Quality Scale (GSQS, Study 2), sleep diary (Study 2) and a short questionnaire to assess subjective sleep quality of the night prior to the cognitive assessment (Study 1). These separate measures capture somewhat different aspects of self-reported sleep quality and thus provide a detailed picture and a more reliable assessment of the participants' subjective sleep quality. In all three studies, cognitive performance was tested not only in working memory and executive functions, but also in several sub-processes of procedural learning. This approach enabled us to test – within the same participants and same experimental designs – whether procedural learning is differentially associated with subjective sleep quality as opposed to working memory and executive functions. To be able to show if there are no associations between the mentioned factors, we calculated Bayes Factors that offers a way of evaluating evidence in favor of the null hypothesis. Finally, we also performed a meta-analysis on the three studies included in the current paper to obtain more persuasive findings as suggested by Goh et al. [17]. To the best of our knowledge, this is the first extensive investigation on the relationship between subjective sleep quality and cognitive functions, covering such a wide range of assessments, in healthy young adults. Our individual studies with frequentist and Bayesian statistical analysis and the meta-analytical approach together provide more reliable results both with direct and conceptual replication.

Materials and Methods

Participants

Participants (all native Hungarians) were selected from a large pool of undergraduate students from Eötvös Loránd University in Budapest. The selection procedure was based on the completion of an online questionnaire assessing mental and physical health status. Respondents reporting current or prior chronic somatic, psychiatric or neurological disorders, or the regular consumption of drugs other than contraceptives were excluded. In addition, individuals reporting the occurrence of any kind of extreme life event (e.g., accident) during the last three months that might have had an impact on their mood, affect and daily rhythms were not included in the study.

In Study 1 there were 47 participants, in Study 2 there were 103 and in Study 3 there were 85. Descriptive characteristics of participants in the three studies are listed in Table 1. All participants provided written informed consent and received course credits for taking part in one of the studies. The studies were approved by the research ethics committee of Eötvös Loránd University, Budapest, Hungary (201410, 2016/209). The study was conducted in accordance with the Declaration of Helsinki.

Table 1. Descriptive characteristics of participants

Study	Number	Age	Sex	Education
Study 1	47	21.38 (1.79)	10M/37F	14.36 (1.58)
Study 2	103	21.62 (2.00)	30M/73F	14.50 (1.74)
Study 3	85	20.99 (1.59)	23M/62F	14.28 (1.60)

All numbers in parentheses show standard deviation. M - male, F - female

Study design

Three separate studies on the association of subjective sleep quality (assessed by sleep questionnaires) and procedural learning (measured with ASRT) and other cognitive functions, such as working memory and executive functions (see Table 2) was conducted. Importantly, the Alternating Serial Reaction Time (ASRT) task used for characterizing procedural learning is designed to capture several sub-processes within this domain. Namely, higher-order sequence learning (i.e., acquisition of order-based information) and statistical learning (i.e., acquisition of frequency-based information) can be separately assessed, and differentiated from other, more general changes in performance (often referred to as general skill improvements), which are not related to the acquisition of order- or frequency-information [18]. For more details, see the Procedural learning subsection of Cognitive performance assessments.

In **Study 1**, subjective sleep quality was measured by the AIS and the short version of PSQI (see Pittsburg Sleep Quality Index subsection of Subjective sleep quality assessments) to characterize participants' general sleep quality over a longer time period (i.e., one month prior to the cognitive testing). In addition, the sleep quality of the night prior to cognitive testing was also assessed by a short lab-designed questionnaire. Cognitive performance was tested in procedural learning (measured by the ASRT task), short-term and working memory (measured by the Digit Span and Counting Span task) and executive functions (measured by the Wisconsin Card Sorting Task, WCST).

In **Study 2**, we included a broader assessment of sleep quality: in addition to the AIS and short version of PSQI for characterizing general sleep quality over a longer time period, a sleep diary was administered to assess sleep quality more precisely (and not retrospectively, but with an ongoing evaluation) over a one to two weeks period prior to the cognitive assessment. GSQS was used to characterize sleep quality of the night prior to the cognitive assessment. Cognitive performance was tested using the same tasks as in Study 1 (that the timing parameters of the ASRT task were slightly modified, for more details see Procedural learning subsection of the Cognitive performance assessments).

In **Study 3**, subjective sleep quality was assessed by the AIS and the shorter version of PSQI. Cognitive performance was assessed in procedural learning (by the ASRT task), working memory (by the Counting Span task), and executive functions (by the WCST). The timing parameters of the ASRT task was modified compared to Study 1 and 2 (see Procedural learning subsection of the Cognitive performance assessments).

In all three studies, AIS and PSQI sleep quality questionnaires were administered online, while the other sleep questionnaires (except for the sleep diary) and the tasks assessing cognitive performance were administered in a single session in the lab. The sleep diary in Study 2 was given to the participants 1 to 2 weeks prior to the cognitive assessment.

Table 2. Tests assessing subjective sleep quality and cognitive performance across the three studies

	Study 1	**Study 2**	**Study 3**
Athens Insomnia Scale (sleep quality in the last one month prior to the cognitive assessment)	**X**	**X**	**X**

Short Pittsburgh Sleep Quality Index (sleep quality in the last one month prior to the cognitive assessment)	**X**	**X**	**X**
Sleep diary (sleep quality of 1-2 weeks prior to the cognitive assessment)		X	
Groningen Sleep Quality Scale (sleep quality of the night prior to the cognitive assessment)		X	
Questionnaire about previous night's sleep quality	X		
Explicit Alternating Serial Reaction Time Task (procedural learning)	**X**	**X**	**X**
Digit Span Task (verbal short-term memory)	X	X	
Counting Span Task (working memory)	**X**	**X**	**X**
Wisconsin Card Sorting Task (executive functions)	**X**	**X**	**X**

The meta-analysis was conducted on the measures which were included in all three studies (marked in bold)

Cognitive performance assessments

Procedural learning - In all three studies, procedural learning performance was measured by the explicit version of the Alternating Serial Reaction Time (ASRT) task (Fig 1, Nemeth, Janacsek & Fiser, 2013). In this task, a stimulus (a dog's head, or a penguin) appeared in one of four horizontally arranged empty circles on the screen, and participants had to press the corresponding button of a Chronos response box (Psychology Software Tools, INC) in Study 1 and Study 2, and a special keyboard with four heightened keys (Z, C, B, and M on a QWERTY keyboard) in Study 3 when the stimulus occurred. The appearance of stimuli followed a predetermined alternating sequence order, such that every second element was part of the sequence and every second element was randomly selected: the dog stimulus always corresponded to sequence elements, and the penguin stimulus indicated random elements (Fig 1A). Participants were informed about this underlying structure of the sequence, and their attention was drawn to the alternation of sequence and random elements by the different visual cues (i.e., dogs vs. penguins). Participants were instructed to respond as quickly and accurately as they could, and also to find the hidden pattern defined by the dog in order to improve their performance.

Figure 1. **Schematic diagram of the procedural learning (ASRT) task design.** A) In this task, the appearance of stimuli is based on a predetermined sequence order, in which pattern and random elements alternate (e.g., 2r4r3r1r, where numbers correspond to the four locations on the screen and the 'r' represents randomly chosen locations). The pattern and random elements are cued differently: the dog stimulus always corresponded to pattern elements, and the penguin stimulus indicated random elements. B) Numbers (corresponding to the locations on the screen) in blue represent elements of the pattern trials (e.g., appearing in the sequential order 2, 4, 3, 1 throughout the task), which were alternating with random elements (green). Because of this alternating structure, some runs of three consecutive trials (triplets) occur more frequently than others (high- vs. low-frequency triplets). For each element, we determined whether it was the last element of a high-frequency triplet (one example in black frame) or low-frequency triplet (examples in magenta frames). C) *Triplet learning* (see text) was calculated as the difference in responses for the last elements of high-frequency triplets (irrespective of random or pattern position) compared to the last elements of low-frequency triplets. *Statistical learning* was assessed by comparing the responses for those *random* elements that were the last elements of high-frequency triplets vs. those that were the lasts of a low-frequency triplets (right column). *Higher-order sequence learning* was assessed as a difference between responses for pattern elements (which are always high-frequency triplets) vs. random-high frequency triplet elements (top row).

The task was presented in blocks with 85 stimuli. A block started with five random stimuli for practice purposes, followed by an 8-element alternating sequence that was repeated ten times. The alternating sequence was composed of fixed sequence (pattern) and random elements (e.g., 2-R-4-R-3-R-1-R, where each number represents one of the four circles on the screen and “R” represents a randomly selected circle out of the four possible ones). The timing of the stimulus differed in the three studies. In Study 1, the stimulus remained on the screen until the participant pressed the correct response button, and the next stimulus was presented 250 ms following the previous response. In Study 2, the stimulus remained on the screen until the participant pressed the correct response button, and the next stimulus was presented 120 ms following the previous response. In Study 3, the stimulus remained on the screen for 580 ms, the participant was asked to respond within this time window, and the next stimulus was presented 120 ms following the previous stimulus. Thus, the task was self-paced with different response-to-stimulus intervals (RSI) in Study 1 and 2, while it was fix-paced (inter-stimulus interval, ISI, of 700 ms) in Study 3. The timing parameters of Study 3 was determined based on previous ASRT studies showing that healthy young adults' average RT performance is around 3700-430 ms during the task [18,19]. We used different settings to explore which timing parameters promote better learning performance. It has been suggested that longer RSI/ISI (e.g., 250 as opposed to 120 ms) can lead to better learning performance as participants have more time to process and elaborate the stimuli [20]. Nevertheless, it is also plausible that the shorter the time between subsequent stimuli, the easier to find the association among them, which is essential in the ASRT task to achieve a good learning performance.

In all three studies, the ASRT task consisted of 20 blocks. As one block took approximately 1-1.5 min, the session took approximately 20-25 min. For each participant, one of the six unique permutations of the four possible stimulus positions was selected in a pseudo-random manner, so that the six different sequences were used equally often across participants (Howard and Howard, 1997; Nemeth et al., 2010).

The alternating sequence of the ASRT task forms a sequence structure in which some of the runs of three successive trials (henceforth referred to as triplets) appear more frequently than others. In the above example, triplets such as 2X4, 4X3, 3X1, and 1X2 (X indicates the middle element of the triplet) occur frequently since the first and the third elements can either be a pattern or a random stimulus. However, 3X2 and 4X2 occur less frequently since the first and the third elements can only be a random stimulus. Fig 1B and 1C illustrate this phenomenon with the triplet 2-1-4 occurring more often than other triplets such as 2-1-3, 2-1-1, and 2-1-2. The former triplet types are termed as *high-frequency* triplets, whereas the latter types are termed as *low-frequency* triplets (see Fig 1C and Nemeth et al., 2013). The third element of a high-frequency triplet is highly predictable (with 62.5% probability) based on the first element of the triplet. In contrast, in low-frequency triplets, the predictability of the third element is less predictable (with 12.5 % probability) based on the first element of the triplet. According to this principle, each trial was categorized as either the third element of a high- or a low-frequency triplet.

Additionally, trials are differentiated by the visual cues (dog vs. penguin) indicating whether a pattern or a random stimulus was presented in that given trial. In case of pattern trials, participants can use their explicit knowledge of the sequence to predict that trial. Consequently, we further differentiate the previously defined high-frequency triplets into two categories based on whether the last element of the triplet was a pattern or a random stimulus. This way, the task consists of three trial types: 1) trials that belong to the explicitly cued sequential pattern and, at the same time, appear as the last element of a high-frequency triplet are termed *pattern* trials; 2) trials of random stimuli that appear as the last element of a high-frequency triplet are termed *random high* trials; and 3) trials of random stimuli that appear as the last element of a low-frequency triplet are termed *random low* trials (see the example in Fig 1C).

Previous studies have shown that as people practice the ASRT task, they come to respond more quickly and more accurately to the high-frequency triplets (irrespective of whether it was for a pattern or a random stimulus) compared to low-frequency triplets (always random), revealing *Triplet learning* (Nemeth et al., 2010; Song, Howard & Howard, 2007). *Triplet learning* is measured as the difference in reaction time (RT) and accuracy (ACC) between high- and low-frequency triplets (RTs of low-frequency triplets minus RTs of high-frequency triplets; ACC of high-frequency triplets minus ACC of low-frequency triplets). Thus, greater Triplet learning is defined as faster/more accurate responses to high-frequency triplets compared to low-frequency triplets. Importantly, however, the comparison of RT and ACC of high- vs. low-frequency triplets does not take into account whether the last elements of the high-frequency triplets are pattern or random stimuli and consequently, provides a mixed measure of at least two separate learning processes.

The two key learning processes that can be disentangled in the explicit ASRT task are the so-called *Higher-order sequence learning* and the so-called *Statistical learning* (Fig 1C). *Higher-order sequence learning* is measured as the difference in RTs between random high and pattern trials (RTs for random high trials minus RTs for pattern trials; ACC for pattern trials minus ACC for random high trials). These trials share the same statistical properties (both correspond to the third element of high-frequency triplets) but have different sequence properties (i.e., pattern vs. random trials). Thus, greater Higher-order sequence learning is defined as faster/more accurate responses to pattern trials compared to random high trials. This learning measure thus can reflect the knowledge about the alternating sequential structure that the participants explicitly acquired during the task.

Statistical learning is assessed by comparing the responses for those random trials that were the last elements of a high-frequency triplet vs. those that were the last elements of a low-frequency triplet (RTs for random low trials minus RTs for random high trials; ACC for random high trials minus ACC for random low trials). These trials share the same sequence properties (both are random) but differ in statistical properties (i.e., they correspond to the third element of a high- or a low-frequency triplet). Hence, faster responses to random high compared to random low trials yields greater Statistical learning. While Higher-order sequence learning quantifies the acquisition of the sequential pattern, Statistical learning captures purely frequency-based learning (Nemeth et al., 2013). Based on previous findings, the cueing of pattern and random stimuli is necessary to promote Higher-order sequence learning, otherwise, it occurs more slowly, and cannot be acquired during a single session [13,18].

Additionally, more general changes in RT and ACC performance can be measured in the ASRT task. These changes occur similarly for all trial types, thus are not related to acquiring the sequential or statistical structure embedded in the stimulus stream. Instead, these general changes indicate general skill improvements, such as more efficient visuo-motor and motor-motor coordination as the task progresses [21], combined with potential fatigue effects that can accumulate during practice [22]. General skill improvements in terms of RT is assessed as the difference of speed in the beginning and at the end of the task (RTs of the first five blocks minus RTs of the last five blocks, see also Statistical analysis). Similarly, general changes can be quantified in ACC between the beginning and the end of the task (ACC of the first five blocks minus ACC of the last five blocks).

In our study, we first report the Triplet learning results because this has been the most common analysis method in the ASRT studies and thus it enables to directly compare our results with those of previous studies. Next, we report Higher-order sequence learning and Statistical learning measures to obtain a more detailed picture of the underlying processes within procedural learning. Finally, we report average RTs and ACCs and their change from the beginning to the end of the end of the task to test whether these more general aspects of performance have a differential association pattern with sleep compared to the learning scores.

Verbal short-term memory **-** The Digit Span Task [23] was used to measure verbal short-term memory (STM). In this task, participants are asked to repeat series of numbers previously read by the experimenter. The series to be repeated are increasingly longer during the task (starting from three-items series to nine-item series). A maximum of four trials was presented at each series length. If the first three trials at a particular sequence length were correctly recalled, the series length was increased by one. The maximum number of digits (i.e., series length) recalled correctly three times provided the measure of the participant's digit span capacity.

Working memory - The Counting Span task [24-27] was used to assess working memory (WM) performance. The task consisted of three series. In each series, each trial included three to nine blue circles as targets, one to nine blue squares, and one to five yellow circles as distractors on a grey background. Participants counted aloud the number of blue circles in each trial, and when finished with counting, they repeated the total number. When presented with a recall cue, participants recalled each total from the preceding set of trials, in the order in which they appeared. The number of presented trials (i.e., set) ranged from two to six. A participant’s counting span capacity was calculated as the average of the highest set sizes of the three series at which the participant was able to recall the totals in the correct serial order.

Executive functions - The Wisconsin Card Sorting Test (WCST; Heaton, 1981; Grant and Berg, 1985) was used to assess executive functions. In this task, participants are asked to find out a sorting rule for cards based on the feedback they receive for their card-sorting choices. During the task, there are four decks on the screen with symbols on them, which differ in three features: number, shape, and color. On the bottom of the screen, a stimulus card appears, and participants have to match this card to one of the decks (based on a sorting rule of their choice). After the choice, participants receive a feedback whether the choice was correct or not. Based on the feedback, participants have to find out the correct sorting rule. In each trial, only one sorting rule is correct (e.g., number, shape or color), and the rule changes several times during the task, allowing to measure adaptation to changing rules. The outcome measure of the task is the number of *perseverative errors*, which shows the inability to change the behavior despite feedback, so the higher values of this measure indicate weaker executive functions.

Subjective sleep quality assessments

***Athens Insomnia Scale* -** The Athens Insomnia Scale (AIS; [8]) was administered in all three studies. AIS is a self-reported questionnaire assessing general sleep quality (over a one month time period), and consists of eight items; the first five items assess difficulty with falling asleep, awakening during the night, early morning awakening, total sleep time, and overall quality of sleep, while the last 3 items pertain to the sense of well-being, overall functioning and sleepiness during the day. Each item of AIS can be rated from 0 to 3 and the total score ranges from 0 to 24, where higher scores indicate poorer sleep quality.

***Pittsburgh Sleep Quality Index* -** The Pittsburgh Sleep Quality Index (PSQI, Buysse et al.,1988) is one of the most commonly used questionnaires measuring self-reported sleep habits and sleep disturbances over the last month. Here we focused on three components of the questionnaire, which were obtained in all three studies: subjective sleep quality, sleep latency and sleep disturbances. Item 6 (referring to the original coding of PSQI) measured the participant's perceived sleep quality, item 5a indicated sleep latency and items 5b-5j (9 items) showed sleep disturbances. These three components range from 0 to 3 and form a global score that ranges between 0 and 9, a higher score indicating poorer sleep quality. Henceforth we refer to this 11-item long PSQI as PSQI.

***Questionnaire about sleep quality of the previous night* -** In Study 1, subjective sleep quality of the night before the cognitive testing was also measured with three items [29]: sleep duration (in hours, similar to PSQI item 4); sleep latency, assessed as the length of falling asleep (in minutes, similar to PSQI item 2); and subjective sleep quality (on a scale from 1 to 5, similar to PSQI item 6). The evaluation of these items was similar to the typical PSQI scoring (third, second and first components, respectively), therefore the three components range from 0 to 3 and form a global score that ranges between 0 and 9, a higher score indicating poorer sleep quality.

***Groningen Sleep Quality Scale* -** In Study 2, subjective sleep quality of the night before cognitive testing was assessed by the Groningen Sleep Quality Scale (GSQS, Mejiman, deVries-Griever and Vries, 1988; Simor, Köteles, Bódizs and Bárdos 2009), which is a 15-item self-administered questionnaire. Every item is a yes or no question, scoring 0 or 1, thus GSQS scores range from 0 to 14 (the first item is typically not scored), a higher score indicating poorer quality of sleep.

***Sleep diary* -** In Study 2, we also asked participants to keep a sleep diary for 1-2 weeks prior the testing session. In this diary, participants had to mark the time they went to bed, the time they got up, and the hours they spent with sleep during this period. After each night, participants also had to rate how good their sleep was (on a scale from 1 to 5), report how long it took them to fall asleep (in minutes), and how many times they woke up during the night. We evaluated data from sleep diaries similarly to PSQI component scores. The average subjective sleep quality was scored as the first component of PSQI; the average sleep latency as the second component of PSQI, the average time spent with sleep as the third component of PSQI, and the average sleep time divided with the time spent in bed (i.e., sleep efficiency) as the fourth component of PSQI. Altogether, based on the sleep diary, we had 4 component scores, ranging from 0 to 3 and form a global score that ranges between 0 and 12, a higher score indicating poorer quality of sleep.

Statistical analysis

Statistical analyses were carried out with the Statistical Package for the Social Sciences version 22.0 (SPSS, IBM), JASP Version 0.8.2.0 (JASP Team, 2017), and the

Comprehensive Meta-Analysis software Version 3 (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005).

Analysis of the ASRT data - To facilitate data processing and to reduce intra-individual variability, the blocks of ASRT were collapsed into epochs of five blocks, following previous ASRT studies [18,27]. The first epoch contained blocks 1–5, the second epoch contained blocks 6–10, etc. We calculated mean accuracy (ACC) for all responses, and median reaction times (RTs) for correct responses only, separately for pattern, random high and random low trials for each epoch. As in previous ASRT studies (Nemeth et al., 2010; Song et al., 2007), two kinds of low-frequency triplets were eliminated: repetitions (e.g. 222, 333) and trills (e.g. 212, 343). Repetitions and trills are low-frequency for all participants, and people often show pre-existing response tendencies to them (Howard et al., 2004). By eliminating these triplets, we attempted to ensure that differences between high vs. low-frequency triplets emerged due to learning and not to pre-existing response tendencies.

Performance in the ASRT task was analyzed by repeated measures analyses of variance (ANOVA) with median RTs or mean ACCs as the outcome measure, and EPOCH (1st-4th epochs) and TRIAL TYPE (pattern, random high, and random low) as within-subject factors. To evaluate the effect of TRIAL TYPE, and thus to confirm that Higher-order sequence learning (pattern vs. random high difference) and Statistical learning (random high vs. random low difference) occurred, Fisher's LSD post-hoc comparisons were performed. Greenhouse-Geisser epsilon (ε) correction was used if necessary. Original *df* values and corrected *p* values (if applicable) are reported together with partial eta-squared (η_p^2) as a measure of effect size. that for the sake of brevity, we do not report ANOVAs for the mixed measure of Triplet learning (RT/ACC difference in responses to high- vs. low-frequency triplets), which does not clearly differentiate between acquiring the sequential and statistical structure embedded in the task. We conducted these ANOVAs, and Triplet learning occurred in all three studies, both in ACC and RT (significant main effect of TRIPLET: all ps < .001).

***Analysis of the relationship between subjective sleep quality and cognitive performance* -** To test the association between subjective sleep quality and cognitive performance, we calculated Spearman correlations and Bayesian Kendall-tau correlations. Subjective sleep quality was assessed by the total scores of each of the sleep questionnaires used in Study 1-3. Among the measurements assessing cognitive performance, the ASRT task of procedural learning provided us with a rich set of learning/performance scores, which were calculated for ACC and RT data as well. We included in our analyses three ACC/RT learning scores, namely, Higher-order sequence learning, Statistical learning, and Triplet learning. that the latter one was included to gain greater comparability with other ASRT studies. In addition to these learning scores, general performance characteristics (average ACC and average RT during the ASRT task), and performance changes (referred to as general skill improvements, and calculated as the change in ACC/RT performance from the beginning to the end of the task) were also included in the analysis. For working memory performance, the digit span and counting span scores were entered into the analysis. For executive functions, the perseverative errors of the WCST were used in the analysis. Normality of data distribution was violated in case of sleep questionnaire scores (based on the kurtosis and skewness of the data as well as the Kolmogorov-Smirnov test), thus we only used robust correlation methods.

The 'classical' (frequentist) correlational analysis has limitations in that a not significant result cannot be interpreted as evidence for *no relationship* between the analyzed scores. To assess the *amount of evidence* for a potential association or no association between sleep quality scales and cognitive measures, we also calculated the Bayes Factor (BF) for the relevant comparisons. The BF is a statistical technique that helps conclude whether the collected data favors the null-hypothesis (H_0) or the alternative hypothesis (H_1); thus, the BF could be considered as a weight of evidence provided by the data (Wagenmakers et al., 2011). It is an effective mathematical approach to show if there's no association between two measures. In case of Bayesian correlation analyses, $H0$ is the lack of associations between the two measures, and $H1$ states that association exists between the two measures. BFs were calculated using JASP. Here we report BF_{01} values. According to Wagenmakers et al. (2011), BF_{01} values between 1 and 3 indicate anecdotal evidence for $H0$, while values between 3 and 10 indicate substantial evidence for $H0$. Conversely, while values between 1/3 and 1 indicate anecdotal evidence for $H1$, values between 1/10 and 1/3 indicate substantial evidence for $H1$. If the BF is below 1/10, 1/30, or 1/100, it indicates strong, very strong, or extreme evidence for $H1$, respectively. Values around one do not support either $H0$ or $H1$. Thus, Bayes Factor is a valuable tool to provide evidence for no associations between constructs as opposed to frequentists analyses, where no such evidence can be obtained based on non-significant results.

Meta-analysis of the three studies **-** To gain more reliable results, we also conducted a meta-analysis on the association between the sleep questionnaires scores and the cognitive measures in the three studies. We included only those measures in the meta-analysis which were assessed in all three studies, namely, the AIS and the short PSQI as sleep questionnaires and indices from the ASRT, Counting Span, and WCST as cognitive performance measures. Effect size measures were the Spearman correlation coefficients. Overall correlation effect sizes and 95% confidence intervals were computed based on the fixed effects model. This model was chosen because the study design, the socio-economic status, age, and health status of the participants, along with tasks and questionnaires used, were similar in the three studies.

Correction for multiple comparisons – To correct for multiple comparisons in the studies, we calculated False Discovery Rates (FDRs). For each study and for the meta-analysis, we calculated FDRs separately, including all comparisons in the given study.

Results

Procedural learning across the three studies

***Study 1* -** The repeated-measures ANOVA on RT data (Fig 2A) revealed a significant main effect of EPOCH ($F_{3,138} = 33.84$, $p < .0001$, $\eta_p^2 = .42$), such that RTs decreased as the learning progressed indicating general skill improvements. The main effect of TRIAL TYPE was significant as well ($F_{2,92} = 60.89$, $p < .001$, $\eta_p^2 = .57$): post-hoc analysis revealed that responses (averaged across epochs) to pattern trials were faster (M = 340.27 ms) compared to random high (M = 348.70 ms, $p = .002$) and random low trials (M = 366.68 ms, $p < .0001$), and responses to random high trials were faster compared to random low trials ($p < .0001$). The different RTs for the different trial types indicate Higher-order sequence learning (difference in pattern vs. random high trials) and Statistical learning (difference in random low vs. random high trials) occurred during the task. The interaction between EPOCH and TRIAL TYPE was not significant ($F_{6,276} = .89$, $p = .41$, $\eta_p^2 = .02$).

Figure 2. Procedural learning performance. RT for correct responses (A, C, E) and accuracy for all responses (B, D, F) as a function of epoch (1-4) and trial type (pattern, random high- and low-frequency trials) in the ASRT task. The gap between the curves of pattern and random high-frequency trials indicates Higher-order sequence learning, the gap between the curves of random high and low-frequency indicates Statistical learning. Error bars are the standard error of mean.

The repeated-measures ANOVA on accuracy data (Fig 2B) again revealed a significant main effect of EPOCH ($F_{3,138} = 13.08$, $p < .0001$, $\eta_p^2 = .22$): ACC decreased during the task. Similarly, to RTs, there was a significant main effect of TRIAL TYPE ($F_{2,92} = 51.06$, $p < .0001$, $\eta_p^2 = .53$). Post-hoc analysis revealed that responses (averaged across epochs) to pattern trials were more accurate (M = 96%) compared to random high (M = 95%,

p = .08) and random low trials (M = 93%, $p < .0001$), and responses to random high trials were more accurate compared to random low trials ($p < .0001$). This again indicates that Higher-order sequence learning and Statistical learning occurred during the task. The EPOCH x TRIAL TYPE interaction was also significant ($F_{6,276} = 4.15$, $p = .001$, $\eta_p^2 = .08$), the ACC for random low trial type decreased more during the task (3.4% decrease on average) than ACC for pattern (1.2% decrease) or random high trial types (1% decrease).

Study 2 - Similarly to Study 1, the ANOVA for RTs (Fig 2C) revealed a significant main effect of EPOCH ($F_{3,306} = 93.13$, $p < .0001$, $\eta_p^2 = .48$): RTs decreased as the learning progressed. We also found a significant main effect of TRIAL TYPE ($F_{2,204} = 63.39$, $p < .0001$, $\eta_p^2 = .38$). Post-hoc analysis revealed that responses (averaged across epochs) to pattern trials were faster (M = 323.01 ms) compared to random high (M = 351.55 ms, p = .002) and random low trials (M = 370.32, $p < .001$), and responses to random high trials were faster compared to random low trials ($p < .001$), indicating Higher-order sequence learning and Statistical learning occurred during the task. The EPOCH x TRIAL TYPE interaction was also significant ($F_{6,612} = 8.74$, $p < .001$, $\eta_p^2 = .08$), RTs for pattern trials decreased more (58 ms on average) than for random high (31 ms) and random low (28 ms) trials.

Again, in the case of accuracy (Fig 2D), there was a significant main effect of EPOCH ($F_{3,306} = 18.12$, $p < .0001$, $\eta_p^2 = .15$): ACC decreased during the task. There was also a significant main effect of TRIAL TYPE ($F_{2,204} = 97.55$, $p < .0001$, $\eta_p^2 = .49$). Post-hoc analysis revealed that responses (averaged across epochs) to pattern trials were more accurate (M = 96%) compared to random high (M = 95%, $p < .001$) and random low trials (M = 92%, $p < .001$), and responses to random high trials were more accurate compared to random low trials ($p < .001$), indicating Higher-order sequence learning and Statistical learning occurred during the task. The EPOCH x TRIAL TYPE interaction was also significant ($F_{6,612} = 5.47$, p

= .0001, $\eta_p^2 = .05$), the ACC for random low trial type decreased more during the task (3.3%) than ACC for pattern (0.7% decrease) or random high trial (1.3% decrease) types.

Study 3 **-** Similar to the previous studies, in the case of RT data (Fig 2E), there was a significant main effect of EPOCH ($F_{3,252} = 45.60$, $p < .0001$, $\eta_p^2 = .35$): RTs decreased during the task. The main effect of TRIAL TYPE was significant as well ($F_{2,168} = 31.58$, $p < .0001$, $\eta_p^2 = .27$). Post-hoc analysis revealed that responses (averaged across epochs) to pattern trials were faster (M = 367.03 ms) compared to random high (M = 375.92 ms, $p = .008$) and random low trials (M = 388.93, $p < .0001$), and responses to random high trials were faster compared to random low trials ($p < .0001$). The different RTs for the different trial types again indicate that Higher-order sequence learning and Statistical learning occurred during the task. The EPOCH x TRIAL TYPE interaction was also significant ($F_{6,504} = .4.50$, $p < .001$, $\eta_p^2 = .05$), the RTs for pattern and random high trial types decreased more (27 ms and 22 ms respectively) than RTs for random low trials (13 ms).

Again, the ANOVA on accuracy data (Fig 2F) revealed a significant main effect of EPOCH ($F_{3,252} = 33.03$, $p < .0001$, $\eta_p^2 = .28$), such as ACC decreased during the task, and a significant main effect of TRIAL TYPE ($F_{2,168} = 78.97$, $p < .001$, $\eta_p^2 =.49$). Post-hoc analysis revealed that responses (averaged across epochs) to pattern trials were more accurate (M = 84%) compared to random high (M = 82%, $p < .0001$) and random low trials (M = 78%, $p < .0001$), and responses to random high trials were more accurate compared to random low trials ($p < .0001$) indicating that Higher-order sequence learning and Statistical learning occurred during the task. The EPOCH x TRIAL TYPE interaction was not significant ($F_{6,504} = .91$, $p = .47$, $\eta_p^2 = .01$).

Short-term/Working memory and executive function performance in the three studies

The verbal short-term memory and working memory capacity, and executive functions (measured as number of perseverative errors in the WCST task) of participants were in the standard range for their age [23,28]. Mean counting span scores were 3.22 (*SD* = 0.80) in Study 1, 3.65 (*SD* = 1.01) in Study 2, and 3.80 (*SD* = 0.84) in Study 3. These average scores represent a mid-range cognitive performance, as the maximum obtainable score is 6. Mean digit span scores were 6.49 (*SD* = 1.04) in Study 1 and 6.25 (*SD* = 1.16) in Study 2, out of the maximum obtainable score of 9. Mean of the number of perseverative errors were 15.48 (*SD* = 5.27) in Study 1, 14.46 (*SD* = 6.37) in Study 2 and 14.76 (*SD* = 5.65) in Study 3 (no maximum score can be defined in this case).

Subjective sleep questionnaire scores in the three studies

Mean AIS score in the three studies were the following: 4.28 (*SD* = 2.76) in Study 1, 3.41 (*SD* = 2.09) in Study 2 and 4.42 (*SD* = 3.04) in Study 3. Overall, these average scores constitute the lower portion of the AIS scale, as scores can range from 0 to 24, with higher scores indicating poorer sleep quality.

Mean PSQI scores were 3.57 (*SD* = 1.54), 2.54 (*SD* = 1.29), and 3.15 (*SD* = 1.82) in Study 1, Study 2, Study 3, respectively. As the obtainable scores on PSQI range between 0 and 9, these average scores represent the lower half of the PSQI scale, although proportionally are somewhat higher (and thus suggest a slightly poorer sleep quality) compared to the AIS scores.

Mean of the Night Before questionnaire score in Study 1, was 2.15 (*SD* = 1.55), which represents the lower portion of the 0-9 scale. Mean of the GSQS score in Study 2 was 2.86 (*SD* = 2.87) on the scale of 0 to 14. Mean of the Sleep diary score in Study 2 was 1.38 (*SD* = 1.22) on the scale of 0 to 12. Altogether, these scores suggest a relatively good sleep quality in the healthy young adult populations of Study 1-3.

PSQI and AIS scores showed moderate to strong correlations in all three studies (r = .72, p <.001 in Study 1; r = .53, p < .001 in Study 2; r = .50 p < .001 in Study 3). The Night Before questionnaire score did not show associations with AIS and PSQI score in Study 1 (r = -.11, p = .45, r = .04 p = .80, respectively). GSQS score however in Study 2 did show weak correlations with AIS, PSQI and Sleep diary scores (r = .31, p =.001, r = .25, p = .01, r = .37 p < .001, respectively). Sleep diary score also correlated with AIS and PSQI scores (r = .38, p < .001, r = .51 p < .001, respectively)

Associations between subjective sleep quality and cognitive performance

In **Study 1,** both the AIS and the PSQI scores showed a moderate positive correlation with the change in accuracy from the beginning to the end of the ASRT task (r = .44, p = .002; r = .40, p = .005, respectively), such that greater AIS/PSQI scores (which correspond to poorer sleep quality) were associated with larger decrease in accuracy during the task. The correlations were supported by the BFs as well, showing substantial evidence for associations between AIS/PSQI and accuracy change in the ASRT task (r = .29, BF_{01} = .09; r = 29, BF_{01} = .09, respectively). Importantly, however, none of these correlations reached statistical significance after FDR correction for multiple comparisons.

No other correlations reached significance, between either of the AIS, PSQI or the previous night's sleep quality scores and the other cognitive performance indices (all ps > .06, uncorrected). The BFs also supported that there are no associations or there is only anecdotal evidence for associations between the remaining measures (all BF_{01} > 0.59). More specifically, in the case of AIS, BFs showed *substantial* evidence for *no* associations between the AIS scores and cognitive performance measured by the ASRT learning and general skill indices (except for ACC general skill improvements discussed above) and by the Counting

Span task (all $BF_{01} > 4.05$). In case of PSQI, again BFs showed *substantial* evidence for *no* associations between the PSQI scores and cognitive performance measured by most of the ASRT indices (except for average ACC and ACC general skill improvements, discussed above) and by the Counting Span task (all $BF_{01} > 3.15$). For the Night Before questionnaire scores, BFs supported *substantial* evidence for *no* associations with most of the ASRT learning and general skill indices (except for RT Higher-order sequence learning score, and ACC general skill improvements, where BFs suggested anecdotal instead of substantial evidence), and with the perseverative error index of WCST (all $BF_{01} > 3.12$). For further details, see Table 3.

Table 3. Associations between sleep questionnaires and cognitive measures in Study 1

	AIS (N = 47)		**PSQI (N = 47)**		**Night Before (N = 47)**	
	Spearman (frequentist)	Kendall tau (Bayes)	Spearman (frequentist)	Kendall tau (Bayes)	Spearman (frequentist)	Kendall tau (Bayes)
ACC learning indices						
ACC Triplet learning	$r = .05$ $p = .76$	$r = .04$ $BF_{01} = 4.95$	$r = .14$ $p = .36$	$r = .10$ $BF_{01} = 3.15$	$r = -.09$ $p = .54$	$r = -.06$ $BF_{01} = 4.40$
ACC Higher-order sequence learning	$r = -.01$ $p = .93$	$r = .008$ $BF_{01} = 5.27$	$r = .01$ $p = .94$	$r = .01$ $BF_{01} = 5.26$	$r = -.03$ $p = .83$	$r = -.02$ $BF_{01} = 5.17$
ACC Statistical learning	$r = -.01$ $p = .94$	$r = -.02$ $BF_{01} = 5.12$	$r = .08$ $p = .58$	$r = .07$ $BF_{01} = 4.18$	$r = .03$ $p = .85$	$r = .01$ $BF_{01} = 5.25$
RT learning indices						
RT Triplet learning	$r = -.06$ $p = .68$	$r = -.04$ $BF_{01} = 4.93$	$r = -.08$ $p = .61$	$r = -.06$ $BF_{01} = 4.58$	$r = -.14$ $p = .34$	$r = -.11$ $BF_{01} = 3.12$
RT Higher-order sequence learning	$r = -.04$ $p = .81$	$r = -.03$ $BF_{01} = 5.01$	$r = .06$ $p = .70$	$r = .05$ $BF_{01} = 4.75$	$r = -.19$ $p = .20$	$r = -.16$ $BF_{01} = 1.62$
RT Statistical learning	$r = -.04$ $p = .80$	$r = .03$ $BF_{01} = 5.11$	$r = -.05$ $p = .75$	$r = -.03$ $BF_{01} = 5.00$	$r = .03$ $p = .85$	$r = .02$ $BF_{01} = 5.19$
General skill indices						
Average ACC	$r = -.11$ $p = .44$	$r = -.07$ $BF_{01} = 4.05$	$r = -.19$ $p = .19$	$r = -.15$ $BF_{01} = 1.92$	$r = .02$ $p = .90$	$r = .02$ $BF_{01} = 5.20$
ACC general skill improvements	$r = .44$ $p = .002^*$	$r = .29$ $BF_{01} = 0.09$	$r = .40$ $p = .005^*$	$r = .29$ $BF_{01} = 0.09$	$r = -.19$ $p = .21$	$r = -.14$ $BF_{01} = 1.97$
Average RT	$r = .06$ $p = .71$	$r = .04$ $BF_{01} = 4.88$	$r = -.14$ $p = .35$	$r = -.10$ $BF_{01} = 3.37$	$r = .11$ $p = .45$	$r = .08$ $BF_{01} = 3.92$
RT general skill improvements	$r = .05$ $p = .75$	$r = .03$ $BF_{01} = 5.07$	$r = -.04$ $p = .81$	$r = -.03$ $BF_{01} = 5.09$	$r = -.03$ $p = .86$	$r = -.03$ $BF_{01} = 5.12$

WM and EF indices						
Counting Span	$r = -.09$ $p = .56$	$r = -.07$ $\text{BF}_{01} = 4.20$	$r = -.10$ $p = .48$	$r = -.09$ $\text{BF}_{01} = 3.70$	$r = -.19$ $p = .21$	$r = -.14$ $\text{BF}_{01} = 2.09$
Digit Span	$r = -.20$ $p = .18$	$r = -.16$ $\text{BF}_{01} = 1.67$	$r = -.17$ $p = .25$	$r = -.13$ $\text{BF}_{01} = 2.38$	$r = -.23$ $p = .13$	$r = -.19$ $\text{BF}_{01} = 1.03$
WCST – perseverative error	$r = .29$ $p = .06$	$r = .23$ $\text{BF}_{01} = 0.59$	$r = .15$ $p = .36$	$r = .11$ $\text{BF}_{01} = 3.00$	$r = -.07$ $p = .68$	$r = -.06$ $\text{BF}_{01} = 4.41$

* significant at level $p < .05$, strong evidence for the H_0 or H_1 by the Bayes Factor (BF) is shaded grey. BF_{01} values between 1 and 3 indicate anecdotal evidence for H_0, and values between 3 and 10 indicate substantial evidence for H_0. Conversely, values between 1/3 and 1 indicate anecdotal evidence for H_1, and values between 1/10 and 1/3 indicate substantial evidence for H_1. Values around one do not support either H_0 or H_1.

In **Study 2,** the PSQI scores showed a weak negative correlation with the average RTs in ASRT ($r = -.22$, $p = .03$), such that higher PSQI scores (which correspond to poorer sleep quality) are associated with faster RTs. The Bayes correlation analysis also showed substantial evidence for the association of the PSQI score and RTs in ASRT ($r = -.17$, $\text{BF}_{01} = .31$). Importantly, however, the correlation did not reach statistical significance after FDR correction for multiple comparisons. No other correlations reached significance between either of the AIS, PSQI, sleep diary or GSQS scores and the other cognitive performance indices (all *p*s > .07, uncorrected). The BFs also supported that there are no associations between the measures (all $\text{BF}_{01} > 0.98$). More specifically, in the case of the AIS scores, BFs showed *substantial* evidence for *no* associations with all but one (RT Higher-order sequence learning) cognitive performance measures (all $\text{BF}_{01} > 4.12$; anecdotal instead of substantial evidence for the remaining measure). In the case of PSQI scores, BFs showed *substantial* evidence for *no* associations with all RT/ACC learning indices, the average ACC, and with most WM/EF indices (all $\text{BF}_{01} > 3.29$; anecdotal instead of substantial evidence for the remaining measures). In the case of the Sleep diary scores, BFs showed *substantial* evidence for *no* associations with all ACC learning, RT Higher-order sequence learning, average ACC and RT general skill learning indices of ASRT, and with the WM/EF indices (all $\text{BF}_{01} >$

3.16). In the case of the score of the GSQS scores, BFs showed *substantial* evidence for *no* associations will all ACC learning indices, most RT learning, and general skill indices, and with all WM/EF indices (all $BF_{01} > 3.69$; anecdotal instead of substantial evidence only for RT statistical learning and average ACC). For further details, see Table 4.

Table 4. Associations between sleep questionnaires and cognitive measures in Study 2

	AIS (N = 103)		**PSQI (N = 103)**		**Sleep diary (N = 97)**		**GSQS (N = 103)**	
	Spearman (frequentist)	Kendall tau (Bayes)	Spearman (frequentist)	Kendall tau (Bayes)	Spearman (frequentist)	Kendall tau (Bayes)	Spearman (frequentist)	Kendall tau (Bayes)
ACC learning indices								
ACC Triplet learning	$r = -.003$ $p = .98$	$r = -.001$ $BF_{01}=7.74$	$r = -.01$ $p = .92$	$r = -.01$ $BF_{01} = 7.71$	$r = -.03$ $p = .75$	$r = -.02$ $BF_{01} =7.13$	$r = -.02$ $p = .87$	$r = -.02$ $BF_{01}=7.56$
ACC Higher-order sequence learning	$r = .06$ $p = .56$	r = .05 $BF_{01}=6.21$	$r = -.02$ $p = .85$	$r = -.005$ $BF_{01} =7.75$	$r = -.05$ $p = .65$	$r = -.04$ $BF_{01} = 6.50$	$r = -.001$ $p = .99$	$r = -.005$ $BF_{01}=7.76$
ACC Statistical learning	$r = -.04$ $p = .66$	r = -.03 $BF_{01}=7.02$	$r = -.03$ $p = .77$	$r =-.02$ $BF_{01} =7.40$	$r = .04$ $p = .72$	$r = .03$ $BF_{01}=6.94$	$r = .03$ $p = .77$	$r = .02$ $BF_{01}=7.38$
RT learning indices								
RT Triplet learning	$r = .11$ $p = .28$	$r = .08$ $BF_{01}=4.12$	$r = -.12$ $p = .24$	$r = -.09$ $BF_{01}=3.29$	$r = -.14$ $p = .16$	$r = -.11$ $BF_{01}=2.15$	$r = -.12$ $p = .23$	$r = -.09$ $BF_{01}=3.13$
RT Higher-order sequence learning	$r = .16$ $p = .11$	$r = .12$ $BF_{01}=1.69$	$r = -.03$ $p = .73$	$r =-.03$ $BF_{01}=7.11$	$r = .02$ $p = .84$	$r = .02$ $BF_{01}=7.39$	$r = -.04$ $p = .68$	$r = -.03$ $BF_{01}=6.97$
RT Statistical learning	$r = -.07$ $p = .46$	$r = -.06$ $BF_{01}=5.08$	$r = -.10$ $p =.30$	$r = -.08$ $BF_{01}=3.86$	$r = -.17$ $p = .09$	$r = -.14$ $BF_{01}=0.98$	$r = -.15$ $p = .14$	$r = -.11$ $BF_{01}=2.11$
General skill indices								
Average ACC	$r = -.03$ $p = .76$	$r = -.03$ $BF_{01}=7.24$	$r = -.05$ $p = .62$	$r = -.05$ $BF_{01}=6.12$	$r = .01$ $p = .92$	$r = .01$ $BF_{01}=7.52$	$r = .14$ $p = .16$	$r = .11$ $BF_{01}=2.28$
ACC general skill learning	$r = -.09$ $p = .39$	$r = -.04$ $BF_{01}= 6.52$	$r = -.19$ $p = .06$	$r = -.12$ $BF_{01}= 1.41$	$r = -.19$ $p = .07$	$r = -.14$ $BF_{01}= 1.04$	$r = -.01$ $p = .96$	$r = -.001$ $BF_{01}= 7.77$
Average RT	$r = -.11$ $p = .26$	$r = -.08$ $BF_{01}=4.23$	$r = -.22$ $p = .03$*	$r = -.17$ $BF_{01}=.31$	$r = -.16$ $p = .10$	$r =-.12$ $BF_{01}=1.51$	$r = -.12$ $p = .22$	$r =-.08$ $BF_{01}=3.69$
RT general skill learning	$r = .01$ $p = .94$	$r = .01$ $BF_{01}= 7.74$	$r = -.16$ $p = .12$	$r = -.12$ $BF_{01}= 1.68$	$r = -.08$ $p = .44$	$r = -.06$ $BF_{01}= 5.05$	$r = -.08$ $p = .43$	$r = -.06$ $BF_{01}= 5.47$
WM and EF indices								
Counting Span	$r = -.01$ $p = .92$	$r = -.001$ $BF_{01}=7.71$	$r = .06$ $p = .53$	$r = .05$ $BF_{01}=6.19$	$r = -.03$ $p = .76$	$r = -.03$ $BF_{01}=7.07$	$r = -.07$ $p = .47$	$r = -.05$ $BF_{01}=5.61$
Digit Span	$r = .08$ $p = .42$	$r = .06$ $BF_{01}=5.16$	$r = .03$ $p = .78$	$r = .02$ $BF_{01}=7.33$	$r = .07$ $p = .51$	$r = .05$ $BF_{01}=5.58$	$r = -.001$ $p = .99$	$r = .004$ $BF_{01}=7.76$
WCST – perseverative error	$r = -.05$ $p = .61$	$r = -.04$ $BF_{01}=6.50$	$r = -.14$ $p = .15$	$r = -.11$ $BF_{01}=2.20$	$r = -.12$ $p = .23$	$r = -.09$ $BF_{01}=3.16$	$r = -.06$ $p = .55$	$r = -.04$ $BF_{01}=6.41$

* significant at level $p < .05$, strong evidence for the H_0 or H_1 by the Bayes Factor (BF) is shaded grey. BF_{01} values between 1 and 3 indicate anecdotal evidence for H_0, and values between 3 and 10 indicate substantial evidence for H_0. Conversely, values between 1/3 and 1 indicate anecdotal evidence for H_1, and values between 1/10 and 1/3 indicate substantial evidence for H_1. Values around one do not support either H_0 or H_1.

In **Study 3,** no correlations reached significance between either of the AIS or PSQI scores and the cognitive performance indices (all *p*s > .08, uncorrected). The BFs also supported that there are no associations between the measures (all $BF_{01} > 1.22$). For both questionnaires, BFs showed *substantial* evidence for *no* associations with almost all ASRT indices (all $BF_{01} > 3.26$). Anecdotal instead of substantial evidence for no associations was shown only for three measures: ACC Higher-order sequence learning score and average ACC in the case of AIS, and ACC Triplet learning in the case of PSQI. For further details, see Table 5.

Table 5. Associations between sleep questionnaires and cognitive measures in Study 3

	AIS (N = 85)		**PSQI short (N = 85)**	
	Spearman (frequentist)	Kendall tau (Bayes)	Spearman (frequentist)	Kendall tau (Bayes)
ACC learning indices				
ACC Triplet learning	$r = -.13$ $p = .25$	$r = -.09$ $BF_{01}=3.33$	$r = -.19$ $p = .08$	$r = -.14$ $BF_{01}=1.22$
ACC Higher-order sequence learning	$r = -.15$ $p = .16$	$r = -.12$ $BF_{01}= 2.04$	$r = -.12$ $p = .27$	$r = -.10$ $BF_{01}=3.07$
ACC Statistical learning	$r = -.05$ $p = .63$	$r = -.04$ $BF_{01}= 6.06$	$r = -.12$ $p = .27$	$r = -.09$ $BF_{01}=3.53$
RT learning indices				
RT Triplet learning	$r = -.07$ $p = .53$	$r = -.07$ $BF_{01}= 4.81$	$r = -.006$ $p = .95$	$r < .001$ $BF_{01}=7.07$
RT Higher-order sequence learning	$r = .01$ $p = .92$	$r = .006$ $BF_{01}= 7.05$	$r = .04$ $p = .72$	$r = .03$ $BF_{01}=6.55$
RT Statistical learning	$r = -.07$ $p = .51$	$r = -.05$ $BF_{01}= 5.56$	$r = -.03$ $p =.27$	$r = -.03$ $BF_{01}=6.64$
General skill indices				
Average ACC	$r = .15$ $p = .17$	$r = .10$ $BF_{01}= 2.68$	$r = .13$ $p = .25$	$r = .09$ $BF_{01}=3.26$
ACC general skill learning	$r = .12$ $p = .28$	$r = .08$ $BF_{01}= 3.72$	$r = .13$ $p = .25$	$r = .08$ $BF_{01}= 3.72$
RT average	$r = -.13$ $p = .24$	$r = -.09$ $BF_{01}= 3.61$	$r = -.10$ $p = .39$	$r = -.06$ $BF_{01}= 4.93$
RT general skill learning	$r = -.03$ $p = .79$	$r = -.02$ $BF_{01}= 6.87$	$r = -.001$ $p = .99$	$r = -.001$ $BF_{01}= 7.07$
WM and EF indices				
Counting Span	$r = .004$ $p = .97$	$r = -.0005$ $BF_{01}= 7.06$	$r = .03$ $p = .82$	$r = .02$ $BF_{01}=6.83$

WCST – perseverative error	*r* = -.04 *p* = .74	*r* = -.03 BF_{01}=6.49	*r* = .04 *p* = .72	*r* = .03 BF_{01}=6.52

* significant at level p < .05, strong evidence for the H_0 or H_1 by the Bayes Factor (BF) is shaded grey. BF_{01} values between 1 and 3 indicate anecdotal evidence for H_0, and values between 3 and 10 indicate substantial evidence for H_0. Conversely, values between 1/3 and 1 indicate anecdotal evidence for H_1, and values between 1/10 and 1/3 indicate substantial evidence for H_1. Values around one do not support either H_0 or H_1.

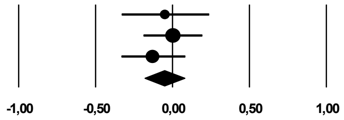
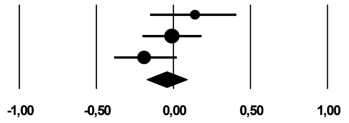
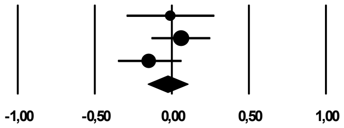
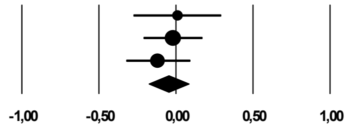
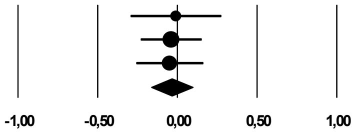
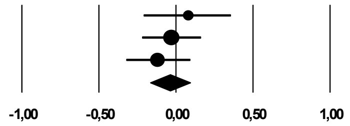
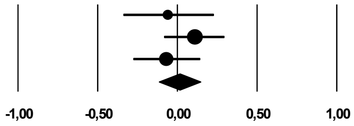
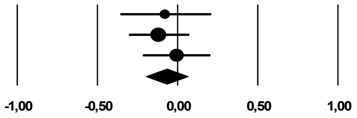
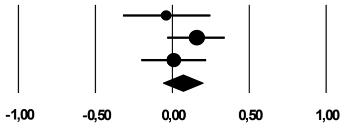
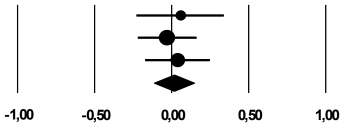
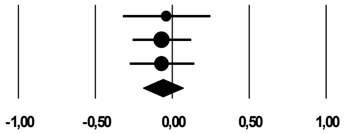
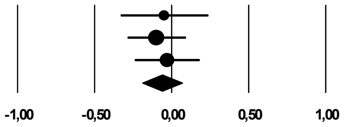
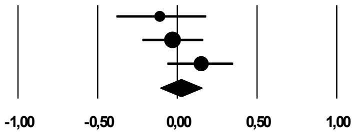
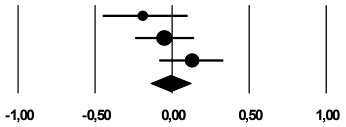
As Spearman and Kendall tau correlation analyses test for linear associations, it would be still possible that non-linear (e.g., quadratic, cubic, etc.) relationship exists between the variables. To rule out this possibility, we also visually inspected all potential associations between subjective sleep quality scores and cognitive measures in all three studies. If this visual inspection had suggested a non-linear relationship, quadratic, cubic or higher-order associations could have been tested in a follow-up analysis. Nevertheless, we did not observe any obvious non-linear pattern in any of the comparisons, and consequently, no further tests were conducted.

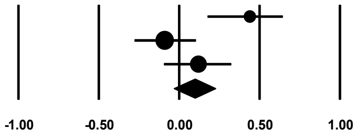
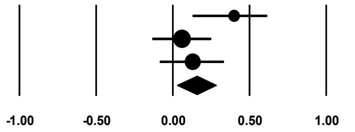
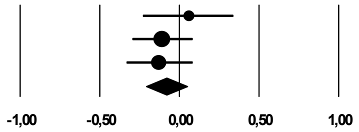
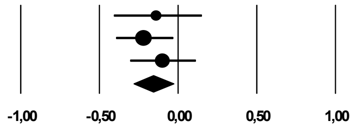
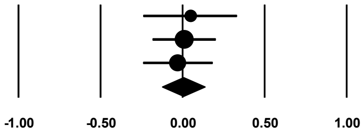
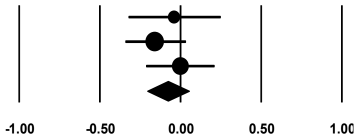
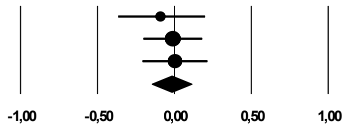
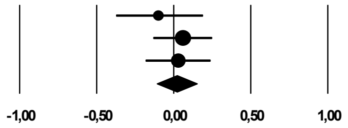
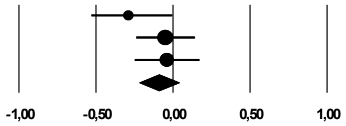
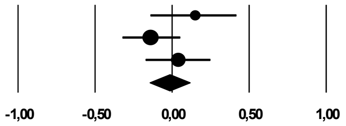
Meta-analysis of the relationship between subjective sleep quality and cognitive performance based on the three studies

To gain more reliable results, we also conducted a meta-analysis on the associations between the sleep questionnaires scores and the cognitive measures in the three studies. We included the AIS and PSQI questionnaire scores as measures of subjective sleep quality, and the Counting Span score and the perseverative error index of WCST as measures of cognitive performance in the meta-analysis because all three studies contained these measures.

In the **meta-analysis** of the three studies (with a sample size of 235 participants), the PSQI was weakly associated with the average RTs of ASRT (*r* = -.16 *p* = .01). Furthermore, the PSQI score was also weakly associated with the change in ACC from the beginning to the end of the task (*r* = .16 *p* = .02). Nevertheless, none of these correlations reached statistical significance after FDR correction for multiple comparisons. The other cognitive measures did not correlate with the PSQI score (all *p*s > .29, uncorrected). In the meta-analysis, the AIS did not show any significant association with any cognitive measures (all *p*s > .15, uncorrected). For further details, see Table 6.

Table 6. Associations between sleep questionnaires and cognitive measures – meta-analysis

	AIS		PSQI short	
ACC learning indices				
ACC Triplet learning	*r* = -.06 *p* = .40		*r* = -.05 *p* = .48	
ACC Higher-order sequence learning	*r* = -.03 *p* = .65		*r* = -.05 *p* = .45	
ACC Statistical learning	*r* = -.04 *p* = .57		*r* = -.04 *p* = .53	
RT learning indices				
RT Triplet learning	*r* = .01 *p* = .86		*r* =-.07 *p* = .29	
RT Higher-order sequence learning	*r* = .07 *p* = .31		*r* = .01 *p* = .85	
RT Statistical learning	*r* = -.06 *p* = .33		*r* = -.07 *p* =.33	
General skill indices				
Average ACC	*r* = .02 *p* = .76		*r* = -.01 *p* = .86	

ACC general skill learning	$r = .10$ $p = .15$		$r = .16$ $p = .02$*	
Average RT	$r = -.08$ $p = .20$		$r = -.16$ $p = .01$*	
RT general skill learning	$r = .003$ $p = .96$		$r = -.08$ $p = .23$	
WM and EF indices				
Counting Span	$r = -.02$ $p = .76$		$r = .02$ $p = .79$	
WCST – perseverative error	$r = -.10$ p = .15		$r = -.02$ $p = .78$	

* significant at level $p < .05$. On the figures, the correlation coefficients and 95% confidence intervals are seen in forest plots. The lines correspond to Study 1, Study 2, Study3 and the meta-analysis, respectively. The sizes of the markers are proportional to the study weights.

Discussion

Our aim was to investigate the relationship between subjective sleep quality (assessed by several self-report measures) and performance in various cognitive functions, such as verbal short-term memory, working memory, executive functions, and procedural learning – which has not been investigated in relation with subjective sleep quality before – and to show whether procedural learning is differentially associated with subjective sleep quality as opposed to short-term/working memory and executive functions. To provide more reliable results, we included robust frequentists and Bayesian statistical analysis and a meta-analytical approach as well. We did not find associations between subjective sleep quality and cognitive performance either in the studies analyzed separately or in the meta-analysis. Moreover, the Bayes factors provided substantial evidence for no associations between subjective sleep quality and most measures of procedural learning, or other cognitive performance measure included in our investigation.

None of the procedural learning indices showed associations with subjective sleep quality (supported by Bayes Factors and the meta-analysis). Higher-order sequence learning, Statistical learning, Triplet learning and general skill learning (in terms of both ACC and RT) seems to be independent of reported sleep quality. In the case of procedural learning, its relationship with objective sleep quality is still debated [30-33]. The results so far have been controversial, as some studies have showed associations between various aspects of sleep, such as time spent in REM sleep [30] or time spent in NREM 2 sleep [31] and procedural learning, while others, focusing primarily on patients with sleep disorders, or examining sleep effects in AM-PM/ PM-AM design have not found such associations [32-36]. Here we focused on subjective sleep quality and showed evidence for no association with procedural learning, which is consistent with previous studies showing no relationship between procedural learning performance and objective sleep measures [32-36]. Importantly, there is a great variability across studies in using different tasks or testing different sleep parameters, as well as in other study settings (e.g., population characteristics, time of the day of testing). More systematic investigations and similar settings (with the same tasks and/or same sleep parameters) across studies are needed to clarify the relationship between sleep quality and procedural learning. Moreover, these systematic investigations could also be submitted to a meta-analysis (such as in this paper) to gain more reliable conclusions.

Contrary to our expectations, short-term/working memory and executive functions also did not show association with subjective sleep quality. As presented in the introduction, some studies reported associations between subjective sleep quality and working memory performance [4], executive functions [3] and decision making [5], although other studies also exist that failed to find such associations [6,10]. These studies focused primarily on healthy/disordered elderly or adolescent populations. To the best of our knowledge, our study is the first that investigates cognitive performance in association with subjective sleep quality in a healthy young adult population. A possible explanation for the diverse results is that when cognitive performance peaks in healthy young adults, subjective sleep quality may not have a substantial impact on it, while in other populations, such as in adolescents, older adults, or in various clinical disorders, where cognitive performance has not yet peaked or have declined, subjective sleep quality can have a bigger impact on performance.

It is also worth noting that sleep quality disturbance is more prevalent in adolescent or elderly populations or in clinical disorders. Consequently, variance and extremities in subjective sleep quality could be greater in these populations, while it can remain relatively small in healthy young adults, possibly leading to ceiling effects (such as that worse sleep quality experienced by healthy young adults is still considered as good quality in other populations). Our participants scored relatively low (indicating good sleep quality) on all sleep questionnaires (see in detail in the Methods section) which could further support this statement. On a related note, it is also an important question whether a poorer sleep quality is relatively transient or persists for years or even for decades. It is plausible and would worth to test *systematically* whether, and in which cases, such long-term poor sleep quality has a more detrimental effect on cognitive performance compared to a relatively more recent decline is sleep quality.

Associations between objective sleep quality (measured by actigraph or electroencephalograph) and various aspects of memory or executive functions have been frequently reported before [1,2]. Here we showed that subjective sleep quality is not associated with short-term memory, working memory and executive function performance.

This dissociation suggests that subjective and objective sleep quality, although measuring the same domains, do not necessarily measure the same aspects of sleep quality and sleep disturbances. Kóbor et al. [37] compared a sleep questionnaire (namely, PSQI), and a sleep diary with actigraphy data. According to their results, while the sleep questionnaire and the sleep diary scores moderately correlated, actigraphy data had only weak correlation with both self-reported measures. Guedes et al. [38] showed that the discrepancy of sleep duration quantified by actigraphy or self-reported measures can even be 1-2 hours on average. Objective and subjective assessments of sleep quality, despite the fact that they often carry labels that imply direct relationship or equivalence, may relate to different parameters [39], such as impressions of sleep quality, restedness, or sleep depth do not appear to be strongly correlated with sleep architecture. Furthermore, subjective sleep quality might be represented by a combination of more than one objective sleep parameter. It is also possible that the important parameters of sleep, that contribute to memory consolidation or executive function performance cannot be captured with self-reported instruments. For example, it is often reported (see also above), that spindle activity or time spent in slow wave sleep (SWS), or in REM sleep is essential for memory consolidation [40-42]. These sleep parameters could not be evaluated subjectively. Also, in laboratory sleep examinations, the general subjective sleep quality together with the sleep quality of preceding days of the examination is usually carefully controlled, thus it is possible, that the parameters showed to be important in the associations of objective sleep parameters and cognitive performance during a given night, can only be measured in case of these carefully controlled conditions, when sleep quality in general and in preceding days are good. Thus it is possible that while objective results show how healthy sleep is in connection with cognitive functioning, subjective sleep quality associations reflect more sleep disorder symptoms and how they are connected to cognitive functioning.

Importantly, we found no associations with cognitive performance both for general sleep quality (assessed for a one-month period) and for the previous night's subjective sleep quality. Although we did not include the results of the previous night's sleep quality in the meta-analysis, the Bayes Factors show evidence for no associations between previous night's sleep quality and procedural learning, short-term/working memory or executive function, both in the case of the Night Before and the GSQS questionnaire.

Considering the dissociation between objective and subjective sleep quality, the usage of self-reported tools to measure sleep quality, in general, should be avoided in investigating the relationship between sleep and cognitive performance, or results should be treated carefully in generalizing results to all aspects of sleep quality. We suggest to use these tests rather for screening participants for sleep disorders in studies where controlling for differences in subjective sleep quality and sleep disturbances is relevant, or sleep diaries for the follow-up of changes after an intervention. The usage of these questionnaires should also be avoided for diagnostic purposes, as also suggested by West et al. [43], who attempted to validate sleep questionnaires with PSG in insomniac patients. In line with this consideration, comparative studies have also shown significant discrepancies between subjective and objective measures of sleep pathology [44,45].

Our paper has some limitations. In meta-analyses, there is no conventional method to correct for multiple meta-analyses conducted on various measures obtained from the same subjects; to the best of our knowledge, this is the first study in which several meta-analyses had been conducted on the data of the same participants. However, we did use multiple comparison correction in case of the meta-analysis as well. Also, even though we included a wide range of cognitive performance measure in our study, it is remained to be investigated, whether self-reported sleep quality is associated with other cognitive tests, such as attentional or other executive function tasks. It is also possible (as mentioned before), that investigating populations more susceptible to sleep disturbances could yield different results, and the lack of associations could be specific for healthy young adults who peak in their cognitive performance and possibly sleep quality as well. Furthermore, we could also test if individual differences influence the relationship between subjective sleep quality and cognitive performance (for example interoceptive ability, i.e. how accurately one perceives their own body sensations).

In conclusion, we showed, across three empirical studies, that self-reported sleep quality is not associated with various aspects of procedural learning, short-term and working memory, and executive function in healthy young adults. These findings were supported not only by classical (frequentist) statistical analyses, but also by Bayes factors (that provided *evidence for no associations* between these functions) and by a meta-analysis conducted on the data obtained from the three individual studies. Importantly, however, our findings do not imply that sleep *per se* has no relationship with these cognitive functions, but it emphasizes the difference between self-reported and objective sleep quality. Together with previous research on dissociations between subjective and objective sleep quality, here we outlined various situations where subjective sleep questionnaires may provide valuable information besides or instead of assessing objective sleep parameters (e.g., for screening sleep disorder symptoms, or investigating populations where performance and sleep quality is not at its peak). Nevertheless, careful consideration should be taken in all cases in order to select the best subjective/objective sleep measures depending on the research question. We believe that our approach of systematically testing the relationship between several self-reported sleep questionnaires and a relatively wide range of cognitive functions can inspire future systematic studies on the relationship between subjective/objective sleep parameters and cognition.

Acknowledgements

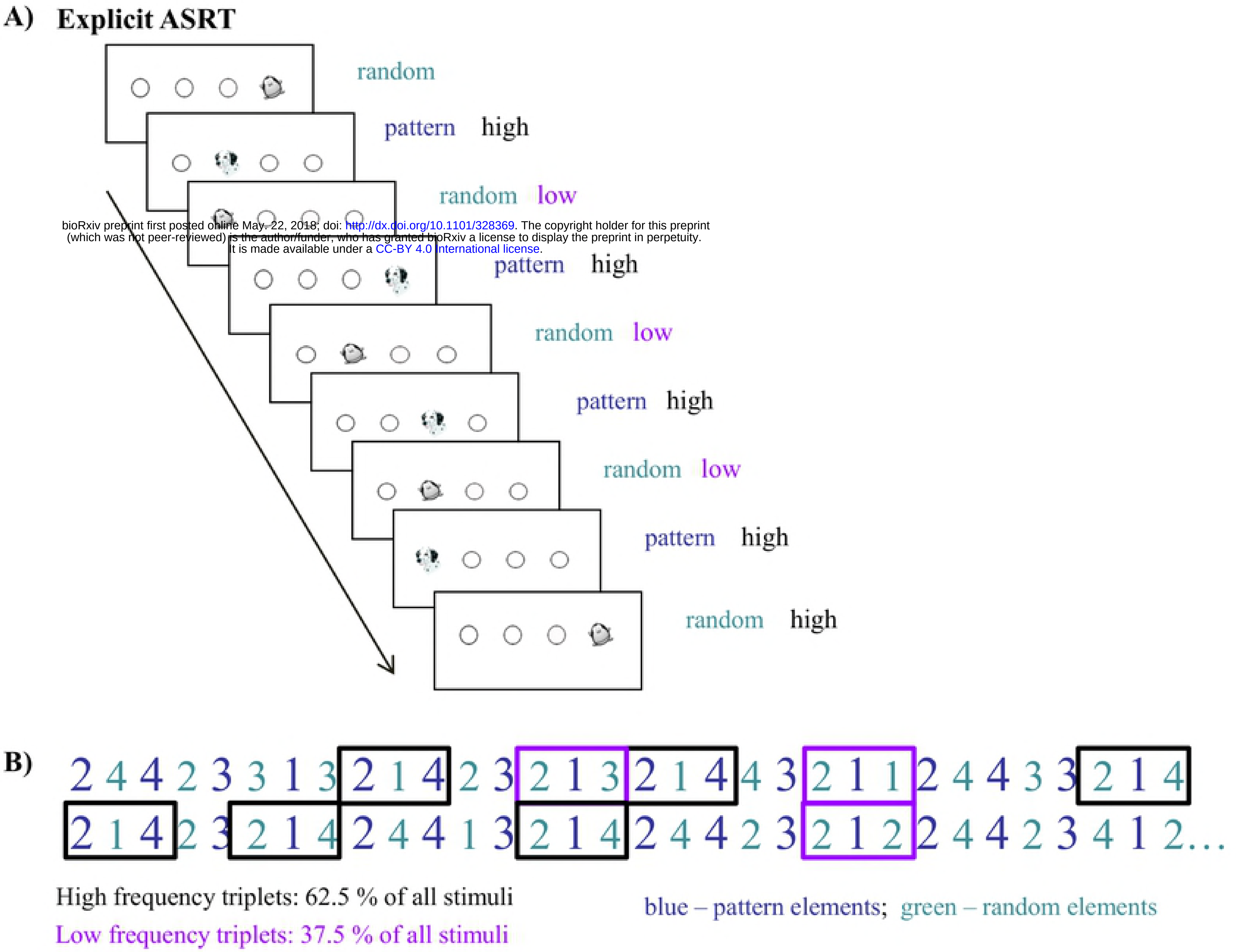
This research was supported by the Research and Technology Innovation Fund, Hungarian Brain Research Program (National Brain Research Program (project 2017-1.2.1-NKP-2017-00002); Hungarian Scientific Research Fund (OTKA PD 124148, PI: KJ); and Janos Bolyai Research Fellowship of the Hungarian Academy of Sciences (to KJ). The authors declare no competing financial interests.

References

1. Diekelmann S, Born J (2010) The memory function of sleep. Nature Reviews Neuroscience 11: 114-126.
2. Jones K, Harrison Y (2001) Frontal lobe function, sleep loss and fragmented sleep. Sleep Medicine Reviews 5: 463-475.
3. Nebes RD, Buysse DJ, Halligan EM, Houck PR, Monk TH (2009) Self-reported sleep quality predicts poor cognitive performance in healthy older adults. The Journals of Gerontology: Series B 64: 180-187.
4. van den Noort M, Struys E, Perriard B, Staudte H, Yeo S, et al. (2016) Schizophrenia and depression: The relation between sleep quality and working memory. Asian journal of psychiatry 24: 73-78.
5. Telzer EH, Fuligni AJ, Lieberman MD, Galván A (2013) The effects of poor quality sleep on brain function and risk taking in adolescence. Neuroimage 71: 275-283.
6. Miyata S, Noda A, Iwamoto K, Kawano N, Okuda M, et al. (2013) Poor sleep quality impairs cognitive performance in older adults. Journal of sleep research 22: 535-541.
7. Buysse D, Reynolds C (1989) The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Research 28: 193-213.
8. Diekelmann S, Biggel S, Rasch B, Born J (2012) Offline consolidation of memory varies with time in slow wave sleep and can be accelerated by cuing memory reactivations. Neurobiology of learning and memory 98: 103-111.
9. Libman E, Fichten CS, Bailes S, Amsel R (2000) Sleep questionnaire versus sleep diary: which measure is better? International Journal of Rehabilitation and Health 5: 205-209.
10. Saint Martin M, Sforza E, Barthélémy JC, Thomas-Anterion C, Roche F (2012) Does subjective sleep affect cognitive function in healthy elderly subjects? The Proof cohort. Sleep medicine 13: 1146-1152.
11. Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso Moyano J, et al. (2001) Interactive memory systems in the human brain. Nature 414: 546-550.
12. Fiser J, Aslin RN (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychological Science 12: 499-504.
13. Howard JH, Jr., Howard DV (1997) Age differences in implicit learning of higher-order dependencies in serial patterns. Psychology and Aging 12: 634-656.
14. Lieberman MD (2000) Intuition: a social cognitive neuroscience approach. Psychological Bulletin 126: 109-137.
15. Poldrack RA, Foerde K (2008) Category learning and the memory systems debate. Neurosci Biobehav Rev 32: 197-205.
16. Pothos EM (2007) Theories of artificial grammar learning. Psychological Bulletin 133: 227-244.

17. Goh JX, Hall JA, Rosenthal R (2016) Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How. Social and Personality Psychology Compass 10: 535-549.
18. Nemeth D, Janacsek K, Fiser J (2013) Age-dependent and coordinated shift in performance between implicit and explicit skill learning. Frontiers in Computational Neuroscience 7.
19. Kóbor A, Janacsek K, Takács Á, Nemeth D (2017) Statistical learning leads to persistent memory: Evidence for one-year consolidation. Scientific Reports 7: 760.
20. Destrebecqz A, Cleeremans A (2003) Temporal effects in sequence learning. In: Jiménez L, editor. Attention and Implicit Learning. Philadelphia, PA: John Benjamins. pp. 181-213.
21. Hallgato E, Győri-Dani D, Pekár J, Janacsek K, Nemeth D (2013) The differential consolidation of perceptual and motor learning in skill acquisition. Cortex 49: 1073-1081.
22. Siegelman N, Bogaerts L, Christiansen MH, Frost R (2017) Towards a theory of individual differences in statistical learning. Phil Trans R Soc B 372: 20160059.
23. Racsmány M, Lukács Á, Németh D, Pléh C (2005) A verbális munkamemória magyar nyelvű vizsgálóeljárásai (Hungarian Diagnostic Tools of Verbal Working Memory Functions). Magyar Pszichológiai Szemle (Hungarian Review of Psychology) 60: 479-506.
24. Case R, Kurland DM, Goldberg J (1982) Operational efficiency and the growth of short-term memory span. Journal of Experimental Child Psychology 33: 386-404.
25. Engle RW, Tuholski SW, Laughlin JE, Conway ARA (1999) Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. Journal of Experimental Psychology 128: 309-331.
26. Conway AR, Kane MJ, Bunting M, Hambrick DZ, Wilhelm O, et al. (2005) Working memory span tasks: A methodological review and user's guide. Psychonomic Bulletin & Review 12: 769-786.
27. Virag M, Janacsek K, Horvath A, Bujdoso Z, Fabo D, et al. (2015) Competition between frontal lobe functions and implicit sequence learning: evidence from the long-term effects of alcohol. Experimental brain research 233: 2081-2089.
28. Heaton RK (1981) A manual for the Wisconsin card sorting test: Western Psycological Services.
29. Nemeth D, Janacsek K, Londe Z, Ullman MT, Howard DV, et al. (2010) Sleep has no critical role in implicit motor sequence learning in young and old adults. Experimental Brain Research 201: 351-358.
30. Fischer S, Hallschmid M, Elsner AL, Born J (2002) Sleep forms memory for finger skills. PNAS 99: 11987-11991.
31. Walker MP, Brakefield T, Morgan A, Hobson JA, Stickgold R (2002) Practice with sleep makes perfect: sleep-dependent motor skill learning. Neuron 35: 205-211.
32. Nemeth D, Csábi E, Janacsek K, Varszegi M, Mari Z (2012) Intact implicit probabilistic sequence learning in obstructive sleep apnea. Journal of Sleep Research 21: 396-401.
33. Wilhelm I, Diekelmann S, Born J (2008) Sleep in children improves memory performance on declarative but not procedural tasks. Learning & Memory 15: 373-377.
34. Csabi E, Benedek P, Janacsek K, Zavecz Z, Katona G, et al. (2015) Declarative and Non-declarative Memory Consolidation in Children with Sleep Disorder. Front Hum Neurosci 9: 709.
35. Csabi E, Varszegi-Schulz M, Janacsek K, Malecek N, Nemeth D (2014) The consolidation of implicit sequence memory in obstructive sleep apnea. PloS one 9: e109010.
36. Simor P, Zavecz Z, Csábi E, Benedek P, Janacsek K, et al. (2017) Delta and theta activity during slow-wave sleep are associated with declarative but not with non-declarative learning in children with sleep-disordered breathing. Sleep Spindles & Cortical Up States 1: 55-66.
37. Kóbor A, Takács Á, Kardos Z, Janacsek K, Horváth K, et al. (2017) ERPs differentiate the sensitivity to statistical probabilities and the learning of sequential structures during procedural learning. Manuscript submitted for publication.
38. Guedes LG, Abreu GdA, Rodrigues DF, Teixeira LR, Luiz RR, et al. (2016) Comparison between self-reported sleep duration and actigraphy among adolescents: gender differences. Revista Brasileira de Epidemiologia 19: 339-347.

39. Armitage R, Trivedi M, Hoffmann R, Rush AJ (1997) Relationship between objective and subjective sleep measures in depressed patients and healthy controls. Depression and anxiety 5: 97-102.
40. Walker MP (2009) The Role of Slow Wave Sleep in Memory Processing. Journal of Clinical Sleep Medicine 5: S20-S26.
41. Siegel JM (2001) The REM sleep-memory consolidation hypothesis. Science 294: 1058-1063.
42. Clemens Z, Fabo D, Halasz P (2005) Overnight verbal memory retention correlates with the number of sleep spindles. Neuroscience 132: 529-535.
43. West G, Vadillo MA, Shanks DR, Hulme C (2017) The procedural learning deficit hypothesis of language learning disorders: we see some problems. Developmental science.
44. Stark-Inbar A, Raza M, Taylor JA, Ivry RB (2016) Individual differences in implicit motor learning: task specificity in sensorimotor adaptation and sequence learning. Journal of neurophysiology 117: 412-428.
45. Dienes Z, Coulton S, Heather N (2018) Using Bayes factors to evaluate evidence for no effect: examples from the SIPS project. Addiction 113: 240-246.



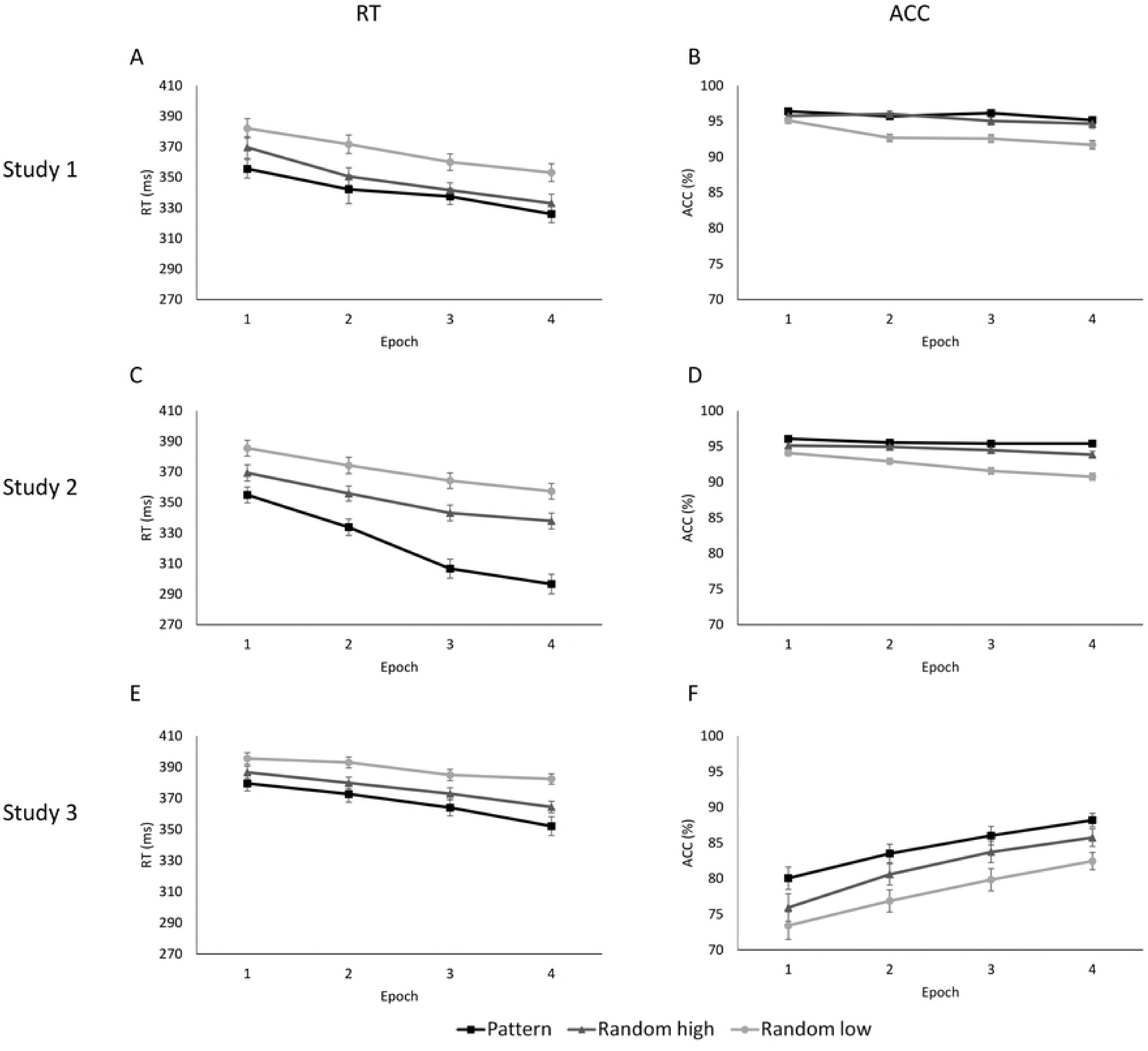
C)

	Structure: 2 – R – **4** (the last event is **pattern**)	Structure: R – 1– **R** (the last event is **random**)
High frequency triplets	2 – 1 – **4** (50%)	2 – 1 – **4** (12.5%)
Low frequency triplets	never occuring (always high)	2 – 1 – **3** (12.5%) 2 – 1 – **1** (12.5%) 2 – 1 – **2** (12.5%)

Higher-order sequence learning:
random high - pattern high

Triplet learning:
Low frequency – high frequency

Statistical learning:
random low - random high


RT
ACC
A
B
C
D
E
F
Study 1
Study 2
Study 3
RT (ms)
410
390
370
350
330
310
290
270
ACC (%)
100
95
90
85
80
75
70
Epoch
1
2
3
4
Pattern
Random high
Random low