Metacognitive accuracy is associated with self-reported sleep outcomes and subjective sleepiness

Rachel A. Jansen (rachel.a.jansen@nasa.gov)
NASA Ames Research Center, Fatigue Countermeasures Laboratory, Moffett Field, CA
Lindsey E. Wade (lindsey.wade@sjsu.edu)
San José State University, Fatigue Countermeasures Laboratory, San José, CA
Cassie J. Hilditch (c.j.hilditch@cqu.edu.au)
San José State University, Fatigue Countermeasures Laboratory, San José, CA
CQUniversity, Appleton Institute, Wayville, SA, Australia
Erin E. Flynn-Evans (erin.e.flynn-evans@nasa.gov)
NASA Ames Research Center, Fatigue Countermeasures Laboratory, Moffett Field, CA

Abstract

Previous work has shown that performance degrades following acute or chronic sleep loss as does the capacity to judge one's own ability to perform. In this online study, we investigate the extent to which individuals are aware of their own ability to perform and what sleep-specific factors may contribute to their judgments of their own ability. In a large online survey-based study with no manipulation of sleep or induced sleep loss, we find that higher self-rated sleepiness and poorer subjective sleep quality are associated with lower accuracy of metacognitive judgments. Self-reported sleep outcomes show an additional relationship with confidence judgments made throughout a task. This provides preliminary evidence that sleepiness and poor sleep quality may affect perceptions of performance in a fraction arithmetic task in addition to actual performance.

Keywords: Metacognition; self-assessment; sleep; math; arithmetic; confidence

Introduction

The ability to assess one's cognitive state (metacognition) is a well-defined research area in cognitive science and education. Contextual elements that might have great impacts on self-assessment ability are often left out of academic studies in these fields.

One contextual element that influences metacognitive ability is sleepiness. Researchers studying metacognition do not typically control for time of day or prior sleep-wake patterns when studying any type of metaawareness. It is well-established that sleepiness caused by sleep loss influences cognitive ability (Van Dongen, Maislin, Mullington, & Dinges, 2003; Mollicone, Van Dongen, Rogers, Banks, & Dinges, 2010) and understanding its relationship to metacognitive ability is important as well.

In laboratory-based studies that have manipulated sleep opportunity, both chronic and acute sleep loss have led to increased errors (Van Dongen et al., 2003) as well as reduced awareness of errors (Boardman et al., 2024), though studies tracking metacognitive monitoring show somewhat mixed results due to study design differences (Boardman et al., 2021). Studies examining the impact of sleep-deprivation on self-assessment of performance have shown sex differences in the degradation of performance and confidence ratings, such as in Rångtell et al. (2019) where sleep deprivation led to a discrepancy between objective and subjective working memory performance in women, but not men.

The American Academy of Sleep Medicine recommends that adults obtain 7–9 hours of sleep per night for adequate cognitive function and behavioral health (Watson et al., 2015), but large-scale studies continue to find that a substantial portion of adults (around 30%) report obtaining less than recommended amounts (Sheehan, Frochen, Walsemann, & Ailshire, 2019; Liu, 2016) and studies using objective measures of sleep find similar patterns (Scott et al., 2024). Without sufficient regular sleep, people are likely to not only perform worse, but also may be poor judges of their own capability.

Going beyond sleep deprivation and its impact on perceived performance, Rahman et al. (2020) manipulated individuals' actual sleep along with their perceived sleep opportunity (e.g., gave them 8 hours to sleep, but changed the clocks to suggest they were only in bed for 5 hours) and found that when participants had a sufficiently long sleep opportunity but believed they slept less, subjective sleepiness ratings increased and actual performance declined, demonstrating that beliefs about one's own sleep outcomes can impact ability potentially beyond actual sleep. This could mean that subjective sleepiness, if grounded in reality, might serve as a useful indicator for someone assessing their current ability.

Nascent research has taken survey-based approaches to studying how general measures of subjective cognition vary with sleep, such as Costa, McCrae, Cowan, and Curtis (2022), who measured the impact of self-reported sleep on both objective and subjective cognition, though their subjective measure was a survey of basic cognitive function rather than questions that probe for perceived performance on specific tests of objective cognition.

For the most part, studies in the sleep field *manipulate* sleep opportunity, typically through acute sleep deprivation or sleep restriction and show that metacognitive ability is not merely an individual trait or domain-dependent, but can also be disrupted by relevant individual experiences (e.g., fatigue, distraction, mood) that affect cognitive ability. These studies typically involve dramatic changes in sleep: a complete loss of a night's sleep (total sleep deprivation) or several hours

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

less of sleep than needed. To further our understanding of the relationship between metacognitive ability, sleep, and sleepiness, in particular everyday variations in obtained sleep, we designed a preliminary study to determine whether and to what extent individuals' metacognition varies under naturalistic conditions (i.e., without manipulating sleep). We test whether individuals who report lower sleep quality the prior night or chronic sleep loss have poorer self-awareness and actual performance or, further, those experiencing slightly more sleepiness might use this sensation as a cue to make more accurate judgments of their own ability.

In this online study, we sought to explore whether there is a direct relationship between sleep and metacognitive ability under naturalistic conditions (i.e., not in a laboratory), with a large, diverse online sample of participants. We employed several different measures of subjective performance at various time points as well as multiple sleep-related surveys to begin to disentangle the specific features of sleep that may influence different dimensions of metacognition.

Methods

Participants

Participants were recruited through Prolific, an online crowd sourcing platform, and required to be at least 18 years old and reside in the United States.

Procedure

The study was administered online using Qualtrics. After providing informed consent, participants completed a brief instructional manipulation. Those who passed were led to a survey of pre-task metacognitive judgments and rated their current level of sleepiness. They next completed a fraction arithmetic task paired with item-level confidence ratings, and then a post-task metacognitive judgments survey. Finally, participants completed a sleep log and a set of sleep-related questionnaires, and ended the study with an entirely optional demographics survey, which included questions about sex, race, and specifics about their sleep environment (e.g., who they live with, noise in their neighborhood).

Stimuli

Fraction arithmetic task. Participants completed 20 sequential multiple-choice fraction arithmetic problems (e.g., $\frac{8}{5} \div \frac{2}{3}$), always administered in the same order. They were instructed not to use calculators or assistive technologies other than pencil and paper and received no feedback throughout the task.

Self-assessed performance. Before being led to the fractions task, participants provided a series of anticipatory metacognitive judgments, including a guess of their expected score on the upcoming task via the prompt, *"You're about to solve a set of 20 fraction arithmetic problems. How many of the 20 questions do you think you will solve correctly?"* (pre-estimate). Immediately following completion of the fractions task, participants responded to the prompt, *"How many of the 20 fraction arithmetic problems you just completed do you think you solved correctly?"* (post-estimate).

Confidence judgments. Prior to each fraction arithmetic item and before seeing the problem, participants rated their confidence on a continuous scale ranging from $0-100$ via the prompt *"You are about to solve a problem. How confident are you that you will solve it correctly?"* These item-level confidence judgments were used to index momentary metacognitve monitoring during task performance and participants answered a total of 21 of these (with one extra final prompt administered after the last problem).

Sleepiness. Current sleepiness was assessed using the Karolinska Sleepiness Scale (KSS), which is a validated and widely used 9-point scale for assessing current levels of sleepiness and has shown some relationship to performance. The scale ranges from: 1 = *"Extremely alert"* to 9 = *"Very sleepy, great effort to keep awake, fighting sleep"* (Åkerstedt, Anund, Axelsson, & Kecklund, 2014).

Sleep log. Participants completed a detailed sleep log inquiring about the prior night's sleep, including subjective sleep quality (1 = *"very good"* to 5 = *"very poor"*), estimated sleep duration, bedtime and wake time, number of minutes needed to fall asleep (sleep onset latency, SOL), time spent in bed after awakening (snooze time), and number and duration of nighttime awakenings (wake after sleep onset; WASO). Participants also indicated their typical nightly sleep duration and bed and wake times in the past month as well as the amount of sleep they believe they need per night to feel fully rested.

Data processing and analysis

Sleep metrics. We generated several metrics based on responses to questions in the sleep log, including: (1) time-in-bed (TIB; duration between self-reported bed and wake times); (2) calculated sleep duration (TIB minus SOL, snooze time, and WASO); (3) acute sleep debt (difference in hours between self-reported sleep needed and self-reported sleep duration); (4) average monthly TIB (duration between self-reported typical bed and wake times); and (5) average chronic sleep debt (difference between sleep needed and self-reported average sleep duration in the past month).

Self-assessment accuracy. We collected several different types of self-assessments throughout the task in order to capture multiple possible metacognitive processes: accuracy, ability to update, and error detection. To illustrate *accuracy* of self-assessed judgments of performance, we calculated the absolute difference between actual and estimated score (both before and after the task).

In order to measure the ability to accurately *update* one's judgments in light of recent evidence, we generated a difference score that numerically illustrates the extent to which participants' estimates improved via the following equation:

$$\Delta = |score - pre| - |score - post| \quad (1)$$

where "score" is an individual's actual score on the fractions task, "pre" is their score estimate made prior to the task, and "post" their estimate made following the task. Positive values denote improved self-assessment (estimates that approach each individual's actual score), negative values indicate worsening self-assessment (the individual's estimated score drifted further from reality after taking the task), and a zero value means no change in distance from actual performance. As an example, if a participant guessed they would answer 10 questions correctly before the task, 15 after the task, and actually achieved a score of 16, their self-assessment difference score (Δ) would come out to 5, meaning they got closer to estimating their actual score by 5 points. Note that a zero value could suggest a lack of updating one's score estimate or equivalently inaccurate estimates in opposite directions.

We also calculated the mean of each participant's 21 confidence ratings as a crude way of representing their *error detection* throughout the task (Boldt & Yeung, 2015).

Analyses. Preliminary analyses were pre-registered on AsPredicted[1]. As planned, we ran linear models predicting self-assessed performance from a sleep metric, controlling for actual score on the fractions task and sex. The sleep and sleepiness metrics used were the previous night's self-reported sleep duration, self-rated sleep quality, average monthly sleep duration in the past month, acute and average chronic sleep debt, and self-rated sleepiness (KSS). We ran these models predicting both score estimates made before and after the fractions task as well as average individual confidence rating. We then performed exploratory analyses on the calculated sleep and self-assessment metrics described above. All analyses were performed using Python (version 3.13.3) and replicated by a second party in R (version 2023.12.1+402).

Results

A total of 544 individuals were recruited to participate, 43 of whom were excluded for: failing instructional manipulation checks (n = 6); not completing the survey (n = 35); or being flagged by Prolific for potential fraud (n = 2). The final analytic sample consisted of 501 participants.

We individually inspected outliers in the sleep log to identify potential participant errors, namely when the difference between self-reported sleep duration and calculated sleep duration was greater than two hours. In seven cases, we discovered participants had made a clear error in their bed or wake time which we fixed (e.g., participant selected 10 AM instead of 10 PM for their bedtime). Four participants made errors in their bed and wake times that could not be reconciled (e.g., bed time at 5 AM, wake at 3 AM) so we excluded their TIB from analyses. Otherwise, we excluded one participant's sleep duration entry, 27 participants' SOLs, three snooze times, one WASO, along with all of these participants' calculated sleep durations given that these metrics all contribute to the calculation (resulting in losing only 3% of calculated sleep duration values). We performed the same investigation for the average monthly TIB and self-reported sleep duration, for which we fixed seven typical bed and wake time entries and excluded an additional seven monthly TIB calculations because of irreconcilable entries.

Of the 501 participants included in data analyses (mean age = 37 ± 13 years), 55.3% were female, 44.1% male and the remaining two declined to specify; 58.3% specified that they were white, 29.7% Black or African American, 2.4% Asian/Asian American, 2.4% Hispanic or Latino. A majority of participants worked regular daytime hours (59.5%) or nighttime hours (28.9%), while the remainder worked either irregular hours, in the evenings, or rotating shiftwork. Most participants reported having no sleep disorder (90.0%) but some had a diagnosed sleep disorder (n = 18, 16 of whom said that their disorder was being managed), or suspected they had a sleep disorder (n = 32) but were not officially diagnosed.

Sleep outcomes

On average, participants reported that they obtained 7 hours and 17 (±1.44) minutes of sleep the previous night and 7 hours and 11 minutes (±1.27) typically on nights in the prior month (see Table 1 for means and distributions of all self-reported sleep outcomes). A total of 20% of participants (n = 101) reported obtaining 6 hours or less of sleep the prior night and 22% (n = 108) on average over the past month. Based on our calculations, the mean prior night's TIB was 7 hours and 49 minutes (±2.04) while the mean monthly TIB was 8 hours and 10 minutes (±2.52), meaning that, on average, they reported obtaining more sleep the prior night compared to the previous month, but spent less time in bed. Calculated prior night's sleep duration was 7 hours and 12 minutes (±2.03), which was very close to their self-reported sleep duration. Participants reported needing 7 hours and 30 minutes (±1.23) of sleep per night to feel rested.

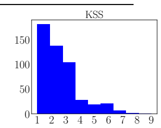
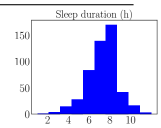
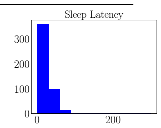
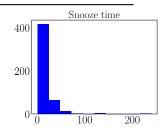
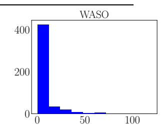
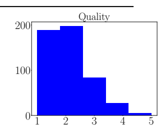
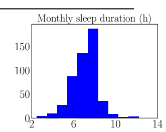
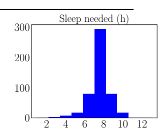
A majority of participants (60%) reported obtaining less sleep the night prior to testing compared to the amount of sleep they said they needed to feel fully rested: on average 1 hour and 1 minute less than needed. The remaining 40%, who reported no acute sleep debt, obtained, on average, 1 hour of sleep beyond what they reportedly needed. Even more participants (70%) experienced chronic sleep debt in the past month (based on the difference in their average sleep duration and sleep needed), averaging 50 minutes of lost sleep per night, while those without self-reported chronic sleep debt obtained 59 minutes of extra sleep per night, which may suggest an inaccuracy in their sleep estimates or incorrect beliefs about the amount of sleep they require.

Self-assessment accuracy

Participants answered between 2 and 20 fraction arithmetic problems correctly (out of 20), though the distribution was

[1] The preregistration is available at https://aspredicted.org/ps49a3.pdf.

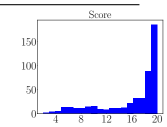
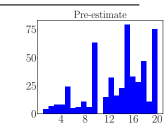
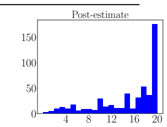
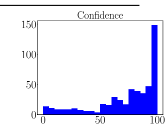
Table 1: Means (standard deviations) and histograms of self-reported sleep outcomes. KSS = Karolinska Sleepiness Scale. SOL = sleep onset latency. WASO = wake after sleep onset.

	M (SD)	
Subjective Sleepiness (KSS)	2.35 (1.51)	
Self-Reported Sleep Duration (h)	7.28 (1.44)	
SOL (min)	20.02 (19.25)	
Snooze Time (min)	14.54 (18.07)	
WASO (min)	6.16 (14.59)	
Sleep quality	1.93 (0.92)	
Sleep Duration in Past Month (h)	7.18 (1.27)	
Sleep Needed (h)	7.50 (1.23)	

negatively skewed ($M_{score} = 16.33 \pm 4.88$, see Table 2). Pre-estimates of scores were on average underestimates ($M_{pre} = 13.69 \pm 4.90$) and score estimates made after completing the fractions task (post-estimates) were still underestimates but higher ($M_{post} = 15.31 \pm 5.56$). Specifically, 347 participants (69%) underestimated their expected score and a total of 235 (47%) underestimated after completing the task.

On average, participants made score estimates prior to the task that were 4.19 ($\pm$3.34) points away from their actual score and 2.16 ($\pm$2.50) points away after the task, suggesting an improvement in self-assessment *accuracy* following the task. In Figure 1, we observe a pattern where the worst performers tended to overestimate their score, while high performers underestimated and score estimates improved from pre to post (neared the red dotted identity line) regardless of actual performance on the fractions task. There were far fewer participants with low scores (see histogram in Table 2), explaining why many of the error bars in Figure 1 are quite diffuse.

Table 2: Means (standard deviations) and histograms of score on the fractions task and metrics of perceived performance.

	M (SD)	
Actual Score (0–20)	16.33 (4.88)	
Pre-Estimate (0–20)	13.69 (4.90)	
Post-Estimate (0–20)	15.31 (5.56)	
Average Confidence (0–100)	73.45 (29.07)	

Preregistered analyses

Score estimates made before and after the task were significantly related to subjective ratings of sleepiness measured via the KSS (pre: $\beta = -0.67$ [-0.91,-0.43], $p < .001$; post: $\beta = -0.42$ [-0.60,-0.23], $p < .001$) and self-rated sleep quality (pre: $\beta = -0.88$ [-1.27,-0.48], $p < .001$; post: $\beta = -0.50$ [-0.80,-0.20], $p < .001$) when controlling for actual score on the fractions task and sex. Note that higher KSS values denote more sleepiness and larger sleep quality ratings refer to worse sleep quality, so in all cases, higher self-reported sleepiness and worse sleep quality were associated with lower self-assessment of performance both before and after the task.

There were no effects of self-reported sleep duration (either the previous night or over the past month) or of acute sleep debt on either of these ratings. However, average chronic sleep debt predicted pre- ($\beta = -0.34$ [-0.62,-0.06], $p = .02$) and post-score estimates ($\beta = -0.23$ [-0.44,-0.02], $p = .03$) with higher chronic sleep loss predicting worse self-rated performance estimates.

Confidence judgments. We performed these same models, but with an individual's mean confidence rating as the dependent variable and found that not only were KSS ($\beta = -3.74$ [-4.88,-2.60], $p < .001$) and sleep quality ($\beta = -5.53$ [-7.42,-3.64], $p < .001$) associated with average confidence ratings, but so were the prior night's self-reported sleep duration ($\beta = 1.36$ [0.14,2.58], $p = .03$), typical sleep duration in the past month ($\beta = 1.92$ [0.55,3.29], $p = .006$), acute sleep debt ($\beta = -1.87$ [-3.08,-0.66], $p = .003$), and average chronic sleep

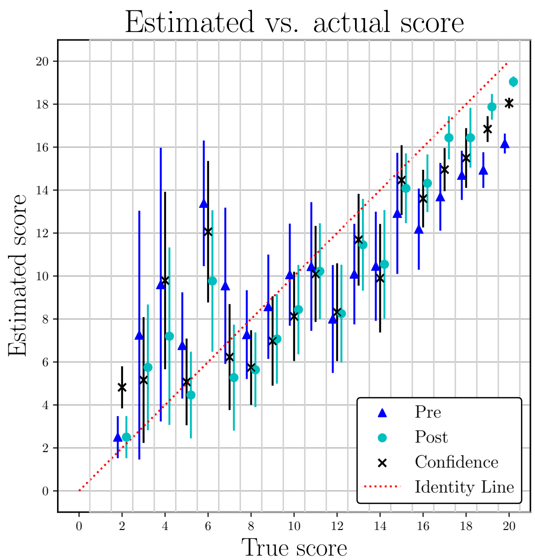


Figure 1: Means of estimated score before completing the task (Pre; blue triangles) and after completing the task (Post; cyan circles) at each true score. Black x's indicate mean confidence judgments at each true score (scaled to be out of 20 rather than 100). Error bars show 95% confidence intervals.

debt ($\beta = -2.50$ [-3.85,-1.15], $p < .001$) with worse self-reported or calculated sleep outcomes predicting lower average confidence ratings.

Calculated sleep metrics. We ran separate models as above predicting each measure of metacognitive ability from our calculated metrics: calculated prior night's sleep duration, prior night's TIB, and past month's typical TIB. Of all the models, only calculated sleep duration showed an association with average confidence ratings ($\beta = 0.97$ [0.06,1.88], $p =$ .037).

Subjective sleepiness

The following exploratory analyses focus on subjective sleepiness ratings given that they were consistently related to each subjective performance metric and because the KSS offers an in-the-moment assessment of sleepiness whereas the time elapsed since an individual woke up and consumption of stimulants might mitigate specific impacts of sleep.

KSS significantly predicted actual score ($\beta = -0.60$ [-0.88,-0.32], $p < .001$), pre-estimates of performance ($\beta = -0.98$ [-1.25,-0.71], $p < .001$), post-estimates ($\beta = -0.97$ [-1.28,-0.66], $p < .001$), and mean confidence ratings ($\beta = -6.03$ [-7.58,-4.49], $p < .001$), such that someone who identified as less sleepy performed better on the fractions task and made higher raw self-assessment judgments (without controlling for their actual score).

Metacognitive updating

The majority of participants updated their estimates to be closer to the true score (n = 299, 60%) and only 79 (16%) deviated further after completing the fractions task. Out of the remaining 123 with a $\Delta = 0$, 106 did not adjust their estimates from pre- to post-task: 37 perfectly estimated their score both before and after the fractions task, 34 overestimated by the same amount both times, and 35 underestimated by the same amount on both occasions.

Those with negative difference scores rated themselves as sleepier ($M_{KSS} = 2.51 \pm 1.74$) and reported receiving less sleep the night prior ($M_{sleep} = 7.07 \pm 1.64$) compared to those with positive difference scores ($M_{KSS} = 2.38 \pm 1.43$; $M_{sleep} = 7.30 \pm 1.37$). In a linear model predicting Δ from KSS, sleep duration, and controlling for actual score and sex, there was a significant effect of KSS such that higher KSS ratings led to higher Δ values ($\beta = 0.23$ [0.05,0.41], $p = .01$) and no effect of self-reported sleep duration. When separately analyzing those who improved in their self-assessments (n=299) from those who worsened (n=79), the relationship was only significant for those with positive Δs ($\beta = 0.21$ [0.002,0.41], $p = .05$) likely because those who were sleepier made larger underestimates of their score from the start and were more accurate after completing the task.

Discussion

In this naturalistic survey-based study, we investigated how and to what extent self-reported sleep outcomes affect metacognitive performance. We found that subjective sleepiness in the moment, self-rated sleep quality, and chronic sleep debt were associated with awareness of performance. All self-reported sleep and sleepiness metrics of interest were related to average confidence, suggesting that confidence ratings made throughout a task might be more reflective of an individual's current cognitive state than global judgments. This is consistent with previous work finding that error detection is specifically affected by sleep loss (Boardman et al., 2024). Additionally, because pre-estimates were made prior to seeing example problems, participants may not have been able to situate their perceived performance in the context of the task at hand. Future analyses will focus on the updating of each individuals' confidence judgments over time.

We applied several methodologies to assess metacognitive ability within a single survey. Calculating simple, interpretable metrics rather than using raw ratings of confidence or subjective general cognitive ability allows researchers to investigate not only whether context influences subjective judgments (e.g., a sleepy person will have lower confidence) but also whether it affects the *accuracy* of these judgments, which epitomizes the meaning of "metacognitive ability." We found that subjective sleepiness is related to how we have operationalized metacognitive *accuracy* as well as *updating* but this may reflect that the sensation of sleepiness is cueing individuals to make lower initial judgments of their ability compared to their less tired counterparts which leads to larger Δ values.

This suggests that sleepiness may act as a useful tool when self-assessing.

Other sleep studies have used subjective sleepiness ratings as the sole measure of subjective experience rather than questions directly targeting perceived performance (e.g., Van Dongen et al., 2003). In sleep-restriction-based protocols, individuals have not shown the same dose-dependent increases in subjective sleepiness as they do decreases in their actual cognitive performance, suggesting that sleepiness might not always be the ideal proxy for self-perceptions of ability. A strength of this study is that we asked for both self-ratings of performance *and* sleepiness and examined their relationship. It may be that subjective sleepiness is better related to performance when sleepiness is generally low (as in our sample) or excessive, but with more mid-range sleepiness, self-rated performance may be more informative.

Better understanding the intricacies of self-assessment ability will impact industries where individuals' fitness for duty could have consequences to safety. In aviation, for example, pilots and air traffic controllers must declare based on their own judgment if they feel fit for duty, but due to the nature of their work, they are often sleep-deprived (e.g., Sallinen et al., 2017). If people experiencing excessive sleepiness are not able to make accurate self-assessments, new methods may be needed for assessing a pilot's fitness for duty.

No objective measures of sleep were collected and thus we must rely on subjective reporting about sleep in this analysis. Therefore, reports of typical sleep habits and prior night's sleep may also be impacted by an individual's cognitive state, but with these data we are unable to assess whether self-reported estimates of sleep duration accurately reflect reality. When asked to self-report sleep duration in other studies, individuals have been shown to overestimate their total sleep time by 20-30 minutes compared to polysomnography (Matthews et al., 2018) which may explain why the participants in this study appeared to sleep more on average than in other studies examining sleep habits (Liu, 2016).

In our study, the mean self-reported sleep duration was higher than we would expect in a representative sample of American adults and within the recommended range (Watson et al., 2015). Only 20% of participants reported obtaining 6 hours or less of sleep the prior night (and 22% on average over the past month) compared to 30% in the typical population as per Sheehan et al. (2019). A mere 6% said they required 6 hours or less of sleep to feel fully functional. These differences between the population recruited and representative groups from other studies may explain why sleep-specific outcomes did not show effects on all measures of metacognitive ability.

Because we did not manipulate sleep or sleep loss, there is very little sleep debt reported overall in the population that we studied. Similarly, the participants' ratings of sleepiness are skewed in a manner that suggests very few participants rated themselves as being on the sleepy end of the scale. Also, the spread of KSS scores is far below the threshold used in safety-managed occupations (≤ 7) to define sleepiness that might affect performance (Åkerstedt et al., 2014). It is interesting that we see any associations with this limited spread of values, but there might be more unexpected relationships that arise at scores higher than 7 on the KSS. Our survey also compared these outcomes between individuals. It is possible that differences in daily sleep need and experience are needed to confirm our findings in a within-subjects cohort to better account for differences between individuals.

Not all participants completed the survey immediately upon waking, which could have introduced recall bias in their reporting of their bed and wake times. Thus there is potential for inaccuracy in the participants' sleep reports. In addition, due to the prevalent use of wearable technology for tracking sleep, it is possible that some participants reported sleep outcomes derived from wearable devices (De Zambotti et al., 2024). However, we did not ask whether participants were using sleep tracking technology. Without an objective measure of sleep specific to the study, we cannot know the extent of sleep misperception among our participants or whether it has any relationship with metacognitive ability. As discussed by Rahman et al. (2020), perceived sleep obtained might differ from actual sleep and this can influence objective performance. Therefore, if a participant slept better or worse than they believed, this may also affect metacognitive judgments. It could be simply that all of these subjective ratings (sleep duration, perceived performance, sleepiness, etc.) are the result of a single cognitive process and may have distinct relationships with objective sleep outcomes, but this will require more involved studies to disentangle. We were careful to place the sleep log in sequence after the fractions task and all metacognitive items so that considerations about sleep would not affect their subjective judgments of performance. However, the KSS was located just before completing the fractions task (though right after the pre-task estimates) deliberately to capture subjective sleepiness at the time of test. This unavoidable methodological choice may have contributed to the consistent relationship KSS scores had with metacognitive beliefs.

Finally, the fraction arithmetic task was relatively simple and very few participants received low scores. With more granularity in both sleep and performance data, there might be more evidence of a clear relationship between sleep and self-ratings of performance.

Future Directions

With these data, we will perform Bayesian modeling on each individual's confidence ratings (as in Jansen, Rafferty, and Griffiths (2020)) to determine distinctions in self-assessment profiles across individuals so we can isolate the metacognitive processes that are associated with sleep and subjective sleepiness (beyond the metrics used in this paper). Future studies will be within-subjects such that we will collect data at several time points to determine if there are individual differences in effects of sleep loss on metacognition and study additional domains.

References

Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of sleep research*, *23*(3), 242–254.

Boardman, J. M., Cross, Z. R., Bravo, M. M., Andrillon, T., Aidman, E., Anderson, C., & Drummond, S. P. (2024). Awareness of errors is reduced by sleep loss. *Psychophysiology*, e14523.

Boardman, J. M., Porcheret, K., Clark, J. W., Andrillon, T., Cai, A. W., Anderson, C., & Drummond, S. P. (2021). The impact of sleep loss on performance monitoring and error-monitoring: a systematic review and meta-analysis. *Sleep medicine reviews*, *58*, 101490.

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*(8), 3478–3484.

Costa, A. N., McCrae, C. S., Cowan, N., & Curtis, A. F. (2022). Paradoxical relationship between subjective and objective cognition: the role of sleep. *Journal of Clinical Sleep Medicine*, *18*(8), 2009–2022.

De Zambotti, M., Goldstein, C., Cook, J., Menghini, L., Altini, M., Cheng, P., & Robillard, R. (2024). State of the science and recommendations for using wearable technology in sleep and circadian research. *Sleep*, *47*(4), zsad325.

Jansen, R., Rafferty, A. N., & Griffiths, T. (2020). A rational model of sequential self-assessment. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

Liu, Y. (2016). Prevalence of healthy sleep duration among adults—united states, 2014. *MMWR. Morbidity and mortality weekly report*, *65*.

Matthews, K. A., Patel, S. R., Pantesco, E. J., Buysse, D. J., Kamarck, T. W., Lee, L., & Hall, M. H. (2018). Similarities and differences in estimates of sleep duration by polysomnography, actigraphy, diary, and self-reported habitual sleep in a community sample. *Sleep health*, *4*(1), 96–103.

Mollicone, D. J., Van Dongen, H., Rogers, N. L., Banks, S., & Dinges, D. F. (2010). Time of day effects on neurobehavioral performance during chronic sleep restriction. *Aviation, space, and environmental medicine*, *81*(8), 735–744.

Rahman, S. A., Rood, D., Trent, N., Solet, J., Langer, E. J., & Lockley, S. W. (2020). Manipulating sleep duration perception changes cognitive performance–an exploratory analysis. *Journal of psychosomatic research*, *132*, 109992.

Rångtell, F. H., Karamchedu, S., Andersson, P., Liethof, L., Olaya Bucaro, M., Lampola, L., . . . Benedict, C. (2019). A single night of sleep loss impairs objective but not subjective working memory performance in a sex-dependent manner. *Journal of Sleep Research*, *28*(1), e12651.

Sallinen, M., Sihvola, M., Puttonen, S., Ketola, K., Tuori, A., Härmä, M., . . . Åkerstedt, T. (2017). Sleep, alertness and alertness management among commercial airline pilots on short-haul and long-haul flights. *Accident Analysis & Prevention*, *98*, 320–329.

Scott, H., Naik, G., Lechat, B., Manners, J., Fitton, J., Nguyen, D. P., . . . others (2024). Are we getting enough sleep? frequent irregular sleep found in an analysis of over 11 million nights of objective in-home sleep data. *Sleep Health*, *10*(1), 91–97.

Sheehan, C. M., Frochen, S. E., Walsemann, K. M., & Ailshire, J. A. (2019). Are us adults reporting less sleep?: Findings from sleep duration trends in the national health interview survey, 2004–2017. *Sleep*, *42*(2), zsy221.

Van Dongen, H. P., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, *26*(2), 117–126.

Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., . . . others (2015). Recommended amount of sleep for a healthy adult: a joint consensus statement of the american academy of sleep medicine and sleep research society. *Journal of Clinical Sleep Medicine*, *11*(6), 591–592.