Online Mobile App Usage as an Indicator of Sleep Behavior and Job Performance

Chunjong Park*, Morelle Arian*, Xin Liu, Leon Sasson†, Jeffrey Kahn†
Shwetak Patel, Alex Mariakakis‡, Tim Althoff
University of Washington, Rise Science Inc.†, University of Toronto‡

ABSTRACT

Sleep is critical to human function, mediating factors like memory, mood, energy, and alertness; therefore, it is commonly conjectured that a good night's sleep is important for job performance. However, both real-world sleep behavior and job performance are difficult to measure at scale. In this work, we demonstrate that people's everyday interactions with online mobile apps can reveal insights into their job performance in real-world contexts. We present an observational study in which we objectively tracked the sleep behavior and job performance of salespeople ($N = 15$) and athletes ($N = 19$) for 18 months, leveraging a mattress sensor and online mobile app to conduct the largest study of this kind to date. We first demonstrate that cumulative sleep measures are significantly correlated with job performance metrics, showing that an hour of daily sleep loss for a week was associated with a 9.0% average reduction in contracts established for salespeople and a 9.5% average reduction in game grade for the athletes. We then investigate the utility of online app interaction time as a passively collectible and scalable performance indicator. We show that app interaction time is correlated with the job performance of the athletes, but not the salespeople. To support that our app-based performance indicator truly captures meaningful variation in psychomotor function as it relates to sleep and is robust against potential confounds, we conducted a second study to evaluate the relationship between sleep behavior and app interaction time in a cohort of 274 participants. Using a generalized additive model to control for per-participant random effects, we demonstrate that participants who lost one hour of daily sleep for a week exhibited average app interaction times that were 5.0% slower. We also find that app interaction time exhibits meaningful chronobiologically consistent correlations with sleep history, time awake, and circadian rhythms. The findings from this work reveal an opportunity for online app developers to generate new insights regarding cognition and productivity.

KEYWORDS

mobile app interaction, interaction time, sleep tracking, sleep behavior, job performance

---

*Both authors contributed equally to this research.

---

1 INTRODUCTION

Sleep is essential to human function, affecting memory [89], energy [20], mood [17], and alertness [3]. The importance of sleep is widely accepted, yet a significant portion of the population does not get sufficient sleep at night [49], and an increasing number of people report experiencing sleep problems [12]. In recent years, sleep tracking has become more commonplace with the introduction of commercially available sleep-tracking technologies like smartphones, smartwatches, mattress sensors, and other devices [50]. Online mobile apps associated with such devices collect and manage sleep data so that users can learn about and improve upon their sleep behavior. In doing so, many people hope to feel more rested and be more productive at their workplace.

The impact of sleep on people's psychomotor function has been widely studied, usually in controlled lab settings. For example, researchers have found that partial sleep deprivation over multiple days can affect people's ability to perform simple tasks like reaction time tasks such as the psychomotor vigilance test (PVT) and mental math [56, 85]. The consequences of sleep deprivation have even been found to be comparable to the cognitive and motor impairments experienced during alcohol intoxication [92]. Although prior literature suggests that poor sleep behavior can impact real-world job performance, this relationship has remained largely unquantified due to the lack of objective measures of both sleep behavior and job performance. Many careers involve a complex combination of cognitive and psychomotor tasks, so it is unclear how contrived tasks like the PVT translate to higher level performance. Furthermore, job performance assessment can require privacy-invasive methods that disrupt a person's work.

Prior literature has leveraged technology interaction patterns as passive and scalable indicators of alertness and other aspects of psychomotor function [1, 36, 59, 66, 67, 69]. In our work, we extend this literature by investigating the relationships between smartphone app-based performance, objective sleep behavior metrics, and objective job performance metrics gathered from two concurrent studies carried out over 18 months. In our first study (**STUDY 1**), we recruited 34 employees from two organizations—a bankruptcy law firm consultancy ($N = 15$) and the National Football League ($N = 19$)—to track their sleep using a mattress sensor, and we compared that data against widely accepted job performance metrics that were provided by their employers. This represents is the largest study to date among those that leverage objective measurements of natural sleep behavior and job performance patterns, with similar studies including less than half the study population and only a single job category [57, 91]. Across 300 nights of sleep with associated job performance metrics, our analyses support the hypothesis that cumulative sleep loss (e.g., sleep debt, sleep history)

is correlated with decreased job performance. We find that one hour of reduced time-in-bed daily for one week was associated with 9.0% fewer contracts established for the average salesperson and a 9.5% grade drop for the average athlete (Section 4.1).

Since job performance metrics can be difficult to capture in practice, we explore the possibility of using timed interactions with the sleep-tracking app as an unobtrusive indicator of broader psychomotor performance. We examined the amount of time participants spent interpreting the information on the app's main screen as an instantiation of an app-based performance metric. We find that app interaction times were correlated with the athletes' game performance (ρ=-0.296, p=0.046), but not the salespeople's performance (ρ=-0.0752, p=0.4106) (Section 4.2).

Although **STUDY 1** is larger than its predecessors, job performance can be extremely diverse and subject to many confounds that are infeasible to track (e.g., external personal issues leading to diminished performance). To corroborate the use of app-based performance as a valid indicator of psychomotor function via its relationship to sleep, we tracked the sleep behavior and app interaction times of 274 individuals (**STUDY 2**). We analyzed this data to determine whether app interaction time is sensitive to known sleep-related influences on psychomotor function while being robust to other individual-level effects like user-specific baselines and smartphone specifications. Across 7,200 tracked nights of sleep with more than 16,000 app interaction events, our analyses reveal that daily variations in app interaction time aligned with constructs in sleep biology, most notably circadian rhythm [3, 16, 25] and sleep inertia [3]. We find that app interaction time was negatively correlated with time-in-bed (ρ=-0.015, p=0.049), sleep history (ρ=-0.055, p=3.9×10^{-11}), and sleep debt (ρ=-0.095, p=5.2×10^{-30}). Furthermore, we demonstrate that participants who lost one hour of daily sleep over the previous week exhibited app interaction times that were 0.5 seconds slower (Section 5.1). In summary, our research investigates the following questions:

RQ.1 Is sleep behavior correlated with job performance? (**STUDY 1**, Section 4.1)
RQ.2 Is app-based performance correlated with job performance? (**STUDY 1**, Section 4.2)
RQ.3 Is app-based performance correlated with sleep behavior? (**STUDY 2**, Section 5.1)

2 RELATED WORK

In this section, we describe prior work on (1) sleep biology, (2) consumer sleep-tracking applications and their effects on users' sleep behavior, (3) the relationship between sleep and performance, and (4) the use of technology to passively infer performance.

2.1 Models in Sleep Biology

Dijk et al. [16, 25] describe sleep biology using an additive two-process model consisting of *circadian rhythm*, the 24-hour biological cycle that occurs in nearly all creatures, and *homeostasis*, the increasing pressure to sleep as one stays awake for longer periods of time. Akerstedt et al. [3] later added a third process: *sleep inertia*, the initial drowsiness that occurs immediately after waking up. Past studies have used two- and three-process models of sleep biology to understand the effects of sleep schedules on mood [35] and athletic

performance [83]. Genetic predispositions and chronotyping (i.e., "morning person" vs. a "night owl") have also been shown to affect sleep behavior [4, 81]. Matchock et al. [62] and Althoff et al. [7] find evidence of significant interaction between circadian rhythms and chronotyping on reaction times. A recent study suggests that female sleep duration is not strongly dependent on menstrual cycles [70].

To characterize the importance of sleep, many studies require that subjects adhere to strict sleep schedules that range from a full night's rest to total sleep deprivation [71]; however, researchers have noted that natural sleep is more commonly characterized by partial sleep deprivation over multiple days, also known as chronic sleep restriction. Regardless of the specifics of one's sleep schedule, researchers have noted the accumulation of sleep debt as an important metric of sleep behavior [26, 27, 41, 85]. Calculating sleep debt requires understanding an individual's sleep need—the optimal amount of daily sleep an individual requires. Sleep need is often measured through a controlled study where a participant is subjected to extended time-in-bed over many days; under these conditions, sleep length typically decays exponentially over time and approaches an asymptote that represents the individual's sleep need [48]. Since sleep need is often difficult to measure in uncontrolled settings, Kitamura et al. [47] propose a method of estimating sleep debt based on one's history of time-in-bed. We utilize Kitamura et al.'s calculation of sleep debt as a cumulative sleep metric in our analyses. Furthermore, we leverage Akerstedt et al.'s three-process model to understand the correlations between sleep debt, job performance, and our app-based performance metric.

2.2 Sleep Sensing Systems

Traditional polysomnography studies utilize expensive sensors like EEGs and EMGs to get fine-grained information about how a person sleeps [42]. Given the growing desire for health-related self-tracking technologies, sleep tracking has become more commonplace. Ko et al. [50] provide a review of consumer sleep sensing technologies. Their review covers sleep-sensing form factors like smartphones [19, 65], smartwatches, wristbands, mattress sensors, and wireless radios [74]. Overall, these technologies are able to gather sleep metrics ranging from sleep duration to disturbance frequency. For instance, Min et al. [65] propose a mobile app that processes seven different sensor streams (e.g., motion, sound, light) to classify a person's sleep state and sleep quality, while Rahman et al. [74] demonstrate that coarse body movements and subtle chest movements from breathing and heartbeats can be detected by measuring the reflections of high-frequency wireless signals.

Beyond exploring new ways of extracting sleep data, another line of research has explored how sleep data should be presented to users and how recommendations should be generated to improve sleep behavior. Bauer et al. [13] show that a recommendation-based peripheral display can serve as a low-effort, yet effective, method for improving awareness of healthy sleep behavior. Daskalova et al. [23] address the creation of personalized recommendations through guided self-experimentation. Their mobile app, SleepCoacher, tracks sleep behavior metrics from the accelerometer and microphone to generate data-driven recommendations; as users engage with these recommendations, SleepCoacher is able to measure whether the

intervention had its intended effect. Daskalova et al. have also explored cohort-based sleep tracking and recommendations [22].

To the best of our knowledge, the aformentioned body of literature has not explored the opportunity of using interactions with a sleep-tracking system as an additional source of information. The act of examining sleep data on a smartphone requires users to exert cognitive load, which itself is tied to sleep behavior. We introduce the notion of app-based performance to investigate whether interactions with a sleep-tracking system can provide insight into sleep behavior or job performance.

2.3 The Effects of Sleep on Performance

In this work, we contextualize large-scale sleep data through job-based and mobile app-based performance measurements [5]. One of the most common tests that have been used for measuring psychomotor function is the psychomotor vigilance test (PVT) [14, 28, 44], during which a person is asked to respond to a visual signal by pressing a button. Researchers have employed a variety of other contrived tasks to measure cognitive and motor performance in relation to sleep. Pilcher and Huffcutt [71] provide a meta-analysis of 19 research studies that examine task performance and mood as a function of sleep restriction. The cognitive tasks Pilcher and Huffcut list in their review include logical reasoning [15], mental math [11], visual search tasks [11], and word memory tasks [63]. The motor tasks they reference include exercise [61, 68], endurance tasks [60], and muscle strength tests [86].

The aforementioned tests have been used to examine the effects of sleep quality on various performance dimensions. Rajdev et al. [75] use the PVT to validate a mathematical model of psychomotor performance based on sleep debt. Ramakrishnan et al. [76] also use the PVT to validate their own phenotype-specific group-average model of psychomotor performance. Lo et al. [56] use a battery of seven cognitive tasks to find that partial sleep deprivation impairs a wide range of cognitive functions, subjective alertness, and mood. Lastly, Killgore et al. [45, 46] study the effects of total sleep deprivation on measures of emotional intelligence, constructive thinking, and decision-making during a gambling task.

Watson [91] provides a literature review on the interaction between sleep and athletic performance, drawing closer to our research questions on sleep and job performance. In one of the most closely related studies to our own **STUDY 1**, Mah et al. [57] examined the effects of sleep on collegiate basketball players. Their study entailed having athletes maintain their typical sleep schedule, after which they underwent a period of sleep extension with a minimum goal of 10 hours in bed each night. The athletes were evaluated using the PVT, subjective scales of sleepiness, and performance metrics specific to basketball practices (e.g., sprint time, shooting percentage). Furthermore, the athletes were asked to rate their own performance during practices and games. Although research like that of Mah et al. strives towards our goal of measuring high-level job performance rather than low-level task performance, their work falls short capturing an objective measurement of in-game performance. In our work, we build on this literature by examining the correlation between sleep behavior and natural, widely accepted job performance metrics collected from the workplaces of athletes and salespeople (**STUDY 1**). Our study represents the largest effort to study this relationship to date, spanning multiple careers
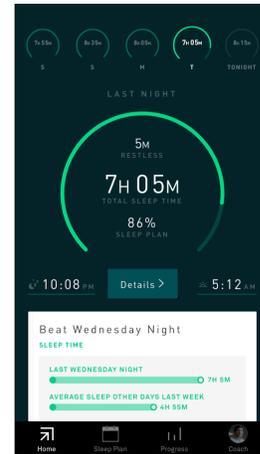
Figure 1: The home screen of Rise Science's sleep-tracking app shows data about the user's most recent night of sleep.

categories and nearly 300 nights of data in a study population that more than doubles prior work [57].

2.4 Interactions as an Indicator of Cognition

Technology interaction patterns have been used as an indicator for understanding different aspects of performance. Regarding smartphones, indicators like app usage and app-specific productivity have been used to estimate alertness [1, 36, 59, 66, 67, 69]. For example, Murnane et al. [66] demonstrate that app usage patterns vary for individuals with different chronotypes and rhythms of alertness. Oulasvirta et al. [67] correlate frequent, short bursts of smartphone interaction (i.e., checking a notification) with inattentiveness. Other researchers have leveraged technology interaction to estimate higher level constructs like stress [32], mood [35, 64], academic performance [90], inebriation [10, 58], and accident risk [6].

Even with just the timing between two interaction events, researchers have been able to assess aspects of a person's cognition. Vizer et al. [87] analyze variation in computer keystroke rate to infer increased stress levels. Althoff et al. [7] track users' typing and click speed on a web search engine as a measure of psychomotor function. Their work shows that keystroke time and click time vary based on sleep duration, circadian rhythms, and the homeostatic process. Inspired by this prior work, we measure app interaction time as a potential non-intrusive indicator of performance (**STUDY 1**) and sleep (**STUDY 2**). Concretely, we demonstrate that this indicator correlates with athletic job performance and meaningfully reflects psychomotor performance variation due to biological functions (i.e., circadian rhythm and sleep inertia).

3 PERFORMANCE AND SLEEP DATA

Both of our observational studies, **STUDY 1** and **STUDY 2**, followed the same protocol. In this section, we first describe the technology that participants used to track and monitor their sleep behavior. We then describe the metrics that we use to quantify sleep behavior, job performance, and app-based performance. We conclude by detailing the procedures that were used to clean the dataset in preparation for analysis.

3.1 Devices

Participants in both **STUDY 1** and **STUDY 2** were recruited as existing users of Rise Science's[1] mobile app. We enrolled participants via targeted recruitment (salespeople and athletes) as well as broader recruitment calls, but participants from all sources were onboarded in a similar fashion. Each participant received a kit consisting of an Emfit QS[2] and a sleep-tracking mobile app[3,4] as shown in Figure 1. The Emfit QS is a highly sensitive pressure sensor that lies underneath the user's mattress (or their preferred side of the mattress when the bed is shared). The sensor uses ballistocardiography to track heart rate, breathing rate, and movement. In past studies, the Emfit QS has been validated against a standard clinical heart rate monitor and polysomnography equipment [37, 51, 78]. Within the sleep-tracking mobile app, participants can access and visualize their own sleep data, view sleep session summaries, create sleep plans, and learn about the importance of sleep.

3.2 Participants and Study Procedure

The data collection period started on May 2017 and ended on December 2018, spanning 592 days. Recruitment happened throughout that period, and participants joined and left the study at their own discretion. In **STUDY 1** (Section 4), participants from the bankruptcy law firm consultancy were enrolled for 225 days, and participants from the NFL teams were enrolled for 450–520 days. In **STUDY 2** (Section 5), participants were enrolled for 80–580 days. Demographic data like age and gender were not collected to minimize intrusion and maintain privacy. However, we know that most NFL players are between 23–27 years old [33], and a public database estimated that 71% of the bankruptcy law firm's employees are millenials[5]. In **STUDY 2**, we expect that most of the participants were under the age of 45 since self-tracking requires active engagement with sleep-technology, which is more common in younger demographics [9].

Special care was taken to avoid coercion during recruitment. To protect participants' privacy, employers were never told who enrolled in the study and were only given aggregated results after the study was done. Participants did not receive explicit instructions from the research team and were free to follow whatever sleep schedule they chose. Participants were also free to use the Emfit QS and sleep-tracking app at will; if they had to travel while participating in the study, they could choose to either bring the Emfit QS with them or leave it behind. The mobile app sent participants notifications, reminders, and recommendations for improving their sleep (e.g., reducing caffeine intake, dimming lights); participants were free to disable these features at any time. Our retrospective data analysis was conducted in accordance with the Institutional Review Board at the University of Washington.

3.3 Sleep Behavior Metrics

The Emfit QS reports the following metrics to describe a single night's rest: bedtime, wake time, sleep midpoint, time-in-bed, and

total sleep duration. Time-in-bed measures how long a person is in their bed, thus only requiring accurate presence detection. Total sleep duration, on the other hand, estimates how long a person is actually asleep in their bed, thus requiring both accurate presence and sleep detection. Because total sleep duration is more susceptible to sensing errors, we exclude it from the analyses reported in this paper and focus on time-in-bed measures, which can be measured with higher accuracy and validity. This choice is common in previous work as well [1, 2, 7, 8, 88]. Nevertheless, sleep duration and time-in-bed were strongly correlated in our dataset ($\rho = 0.85$, $p < 0.001$) and produced comparable results in most cases. Looking beyond a single night's sleep, cumulative metrics across multiple nights can provide further insight into participants' sleep behavior. We use sleep debt [27, 47, 85], the weighted accumulation of sleep loss, as one of those measures. Sleep debt is calculated using the following formula:

$$\sum_{i=1}^{7} -e^{-i/7} * (\text{SleepNeed} - \text{TimeInBed}_i)$$

where i is the number of days in the past. Note that the difference between sleep need and debt is weighted by a decaying exponential with a time constant of 7 days [77], indicating that recent measurements have greater importance. Whenever a participant skipped a day of tracking, we impute the missing time-in-bed value using their average time-in-bed over the past week. Significant imputation happened in only 12.7% of the weeks (see Section 3.6 for details). Sleep need is typically estimated in a controlled laboratory study, making it challenging to estimate sleep debt in the wild. Therefore, we estimate sleep need using the approach proposed by Kitamura et al. [47]. Their approach involves using long nights of sleep to predict the difference between sleep need and habitual sleep (i.e., the average time-in-bed over two weeks) for a minimum of four nights. We also introduce a simplified *sleep history* metric that avoids the notion of sleep need but still captures an aggregate measure of sleep behavior:

$$\frac{1}{\sum_{n=1}^{7} e^{-n/7}} \sum_{i=1}^{7} e^{-i/7} * \text{TimeInBed}_i$$

The calculation of sleep history is normalized such that weights sum to one, making the metric more interpretable as a weighted average of time-in-bed over the past week.

3.4 Job Performance Metrics

For **STUDY 1**, we were able to utilize organizational partnerships to gather the job-specific performance metrics described below:

3.4.1 Performance Metrics for Salespeople. The salespeople who participated in our study work at a bankruptcy law firm consultancy. Their job entails fielding phone calls from potential clients in need of bankruptcy relief and referring those callers to an attorney. The employees collect a fee upon successfully hiring a client, which is the company's primary revenue source. Employees in this company are evaluated on a variety of metrics related to that revenue stream, such as the amount they collect in fees. However, the distribution of fees is highly variable ($250–$1750) and primarily dependent upon the clients rather than the employees themselves. Therefore,

---

[1]https://www.risescience.com/

[2]https://qs.emfit.com/

[3]https://play.google.com/store/apps/details?id=com.risesci.risesciapp

[4]https://apps.apple.com/us/app/rise-science/id1107659850?app=itunes&ign-mpt=uo%3D4

[5]The firm's name is removed to preserve their anonymity.

	Metric	Description
Sleep	Bedtime	Time at which the user got into their bed
	Wake time	Time at which the user got out of their bed
	Midpoint	Midpoint between start and end time
	Time-in-bed	The total time the user spent in bed during a single day including nighttime sleep and naps, regardless of whether they were sleeping
	Sleep debt	Weighted average of difference between sleep need and time-in-bed
	Sleep history	Weighted average of time-in-bed
Job Performance	Number of hires (salespeople)	Number of contracts made after consulting, normalized by the number of hours they work
	Game grade (athletes)	Score of a player's game performance out of 100 assigned through three independent experts
App Usage	Interaction time	Time between opening home screen of app to another screen by user's touch input

Table 1: A summary of the metrics we collect in our dataset through three data streams: (1) sleep metrics through the Emfit QS, (2) job performance through the participants' employers, and (3) app usage through a sleep-tracking mobile app.

we focus on the *number of hires per day* the salespeople were able to establish as their job performance metric. This metric follows a right-sided normal distribution [55] with a mean of 3.80 and a standard deviation of 3.27 hires per day. Although work hours were generally consistent across the company, we normalized the number of hires a salesperson made by the number of hours they worked that day to account for whatever variance remained.

3.4.2 Performance Metrics for Athletes. The athletes who participated in our study play in a professional American football league in the United States. We gather job performance metrics for the athletes' performance during weekly games using Pro Football Focus[6] (PFF). PFF evaluates athletes using the following procedure [72]: two experts score every play the athlete is involved in, a third expert resolves disagreements between those experts, an external group of ex-players and coaches verifies the scores, and then the scores are summed together and normalized to a grade between 0–100. Although PFF is not purely quantitative, the experts can account for in-game context that is lost by purely statistical methods (e.g., injuries, matchups). For this reason, PFF has been used in the past literature for assessing performance in football [18, 29, 73].

In American football, each player has their own unique skill set according to their position; for example, quarterbacks are typically known for their throwing ability and wide receivers are known for their speed and catching ability. The notion of positional specialization makes it difficult to compare athletes across positions in a purely quantitative way, especially since some skills are position-specific. Nevertheless, PFF's method of expert observation and score normalization allows them to produce an *overall game performance* grade that can be used to compare athletes across positions.

3.5 Sleep-Tracking App Usage Metrics

Participants had to interact with a sleep-tracking app in order to examine their sleep summaries, so we leverage these interactions as a novel source of data. We take inspiration from Althoff et al. [7] by using *app interaction time*—the time between two touch events in the app—as an app-based performance metric. App interaction time is not meant to be a direct replacement of the PVT; instead, it serves as a more general measure of cognition by measuring the user's

ability to process information on the app's screen. Interaction speed can be confounded by the content that is shown on the screen. To account for this confound, we restrict our analysis of app interaction time to transitions from the home screen (shown in Figure 1) to other endpoints within or outside of the app.

3.6 Data Filtering and Post-Processing

Sleep behavior, job performance, and app interaction metrics (Table 1) were collected from separate sources at different intervals. Therefore, post-processing was needed to join and collate them.

3.6.1 General Post-Processing. We followed best-practices in preparing mobile app data for analysis [40]. The calculation of time-in-bed included naps, which were either automatically annotated if the user's bedtime or wake time fell in the afternoon (12:00-18:00) or manually annotated by the user. Naps appeared in 9.3% of the nightly sleep metrics (62% automatically tagged vs. 38% manually annotated), contributing an additional 1.22 hours to time-in-bed on average. Sleep events when participants spent more than 16 hours in bed in a single session were attributed to faulty sensing and removed from the dataset. The remaining nights, along with imputed averages for missing values, were used for calculating sleep debt and sleep history. A full week of sleep data was available for calculating 46.9% of the cumulative sleep metrics, meaning that no imputation was needed for them; three or more nights were only missing in 12.7% of the cumulative sleep metrics. When cumulative sleep metrics were calculated without imputation, the standard deviation of the times within the same week was only 1 hour and 10 minutes; this shows that there was not significant variance within a week, justifying the use of a short-term average. For the analyses related to app-based performance, interaction events that were shorter than 0.45 seconds (2.5th-percentile) were excluded since these were likely accidental or automatically generated by the app itself; events longer than 54.83 seconds (97.5th-percentile) were excluded since they were likely indicative of the user engaging in another activity.

3.6.2 Job-Specific Filtering. Job performance data for the salespeople was collected on a daily basis. Therefore, every night of sleep that a salesperson tracked with their Emfit QS was collated with

STUDY 1 Statistics	Salespeople	Athletes
Number of participants	15	19
Total unique days with both sleep-tracking and job performance measurements	118	171
Total unique days with both app interaction and job performance measurements	122	46
Total nights of sleep tracked with app-based performance measure	234	418
Total nights of sleep tracked	834	2,687
Total number of transitions between screens	679	909
Total number of times app was opened	425	691
Nights of sleep tracked per user (avg ± std)	46.33 ± 37.45	133.1 ± 89.92
Time-in-bed in hours (avg ± std)	7.283 ± 2.020	7.308 ± 1.920
Days of app use per user (avg ± std)	28.25 ± 21.06	40.65 ± 50.43

Table 2: Summary statistics for our dataset in STUDY 1 after the filtering described in Section 3.6.

		Sleep Metrics					
		Raw Metrics			Per-Person Z-Normalized Metrics		
		Time-in-Bed	Sleep Debt	Sleep History	Time-in-Bed	Sleep Debt	Sleep History
Job Performance Metrics	NFL Player Game Grades ($N = 19$)	-0.024 (p=0.751)	-0.095 (p=0.218)	-0.029 (p=0.711)	0.086 (p=0.263)	**0.166 (p=0.031)**	**0.179 (p=0.020)**
	Salespeople Hires per Day ($N = 15$)	-0.067 (p=0.469)	**0.218 (p=0.022)**	0.039 (p=0.690)	-0.102 (p=0.283)	0.164 (p=0.088)	-0.047 (p=0.634)

Table 3: Spearman correlation coefficients between sleep behavior and job performance. P-values are provided in parentheses; results with p-value < 0.05 are shown in bold.

the job performance metric from the next day. Aligning the data streams for the athletes was more difficult since they had games on a weekly basis. The athletes also had to travel to games away from their home stadium, leaving larger gaps in their sleep-tracking data. To accommodate these issues, we aligned the weekly PFF grades with the sleep behavior metrics from the most recent tracked night of sleep within the two nights before the relevant game day; if no nights were tracked in that span, the game grade from that week was filtered out.

3.7 Distribution of Job Performance Data

Using D'Agostino's K^2 test [21], we determined that the job performance metrics in our dataset were non-normally distributed (number of hires: K^2=21.37, p=2.3×10^{-5}; game grades: K^2=14.87, p=5.9×10^{-4}). The same holds true for app-based performance (K^2 = 5177, p<1.0×10^{-20}) and app event count (K^2=71.60, p=2.8×10^{-16}). Therefore, we use Spearman's Rank Correlation (ρ) across all correlational analyses throughout this paper.

4 STUDY 1: ATHLETES AND SALESPEOPLE

Our first study investigates **RQ.1** and **RQ.2** within a cohort of 15 salespeople and 19 athletes. Table 2 shows the summary statistics of our dataset after post-processing. The large standard deviations in the various metrics are due to the logistics of our study. Participants were recruited throughout the 18-month-long period, so some people had many more opportunities to use the sleep-tracking tools than others.

4.1 RQ.1: The Relationship Between Sleep Behavior and Job Performance

Using the objective job performance metrics we were able to obtain from our participants' employers, we first examine whether better sleep behavior improves job performance. To the best of our knowledge, our study is the largest to date on this topic without any constraints on how participants slept or went about their daily jobs [56, 57]. The code for all of our analyses can be found in the GitHub repository associated with this project[7].

4.1.1 Analysis Procedure. For this analysis, we calculate correlation coefficients between the job performance metrics and three sleep behavior metrics: time-in-bed, sleep debt, and sleep history. Sleep metrics can vary across individuals due to genetic predisposition and other possible confounds [4, 81], so we repeat the analysis using standardized sleep behavior metrics according the Z-score within each individual's data. Participants who did not track at least 5 nights of sleep were excluded from this analysis to ensure that the data was representative of their typical sleep behavior. The salespeople and athletes contributed data from 118 and 171 nights of sleep with corresponding job performance metrics, respectively.

4.1.2 Results. The correlation coefficients between the sleep behavior and job performance metrics in our dataset are presented in Table 3. The analysis reveals positive, statistically significant correlations in some, but not all, cases. For the salespeople, sleep debt was positively correlated with the number of hires they made (ρ=0.218, p=0.022). For the athletes, normalized sleep history (ρ=0.179, p=0.020) and sleep debt (ρ=0.166, p=0.031) were both positively correlated with game performance. Fewer correlations were found for the salespeople than the athletes, which could be due to the nature

---

[7]Code available at https://github.com/cjpark87/mobile-app-sleep-performance.

(a) Sleep Debt

(b) Normalized Sleep History


(c) Normalized Sleep Debt
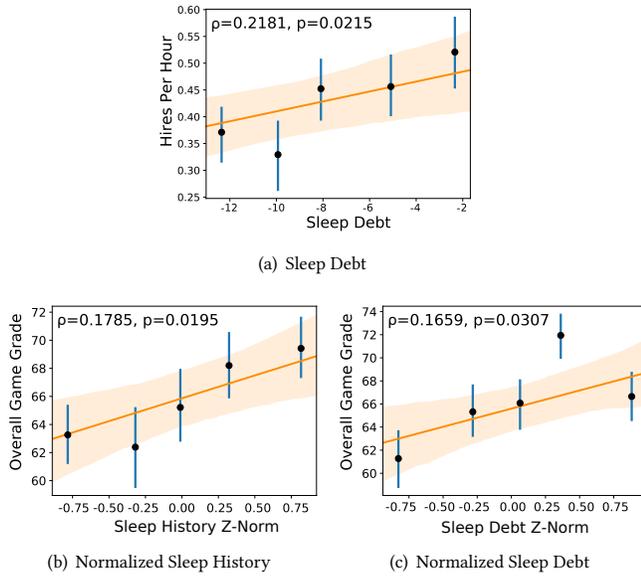
Figure 2: Regression plots showing the effect sizes for the statistically significant results from Table 3. The job performance of both the (a) salespeople and (b+c) athletes is sensitive to cumulative sleep metrics. Throughout the paper, the data are binned into discrete and evenly distributed intervals (quintiles or deciles). Point estimate and error bars represent mean estimate (black) and standard error (blue), respectively. Orange lines represent the best linear regression fit to the raw data along with standard errors (shaded area).

of their jobs. The athletes rely on millisecond-scale reaction times during their games, whereas salespeople do not need to operate at such a rapid pace. These results could imply that careers focused on physical and psychomotor skills may be more strongly affected by sleep behaviors than careers that focus primarily on cognition. The fact that multiple correlations emerged between cumulative sleep behavior metrics and job performance, combined with the lack of such correlations from single-day metrics, suggests that sleep over an extended period has a stronger impact on a person's job performance than a single night of sleep. Additionally, the general increase in correlation coefficients after the sleep behavior metrics were normalized within individuals supports the notion that sleep needs and behaviors vary between individuals.

We further analyze the statistically significant correlations by measuring their effect sizes, which are shown in Figure 2. One hour of sleep debt by the average salesperson was associated with a 2.2% decrease in the number of hires they were able to make. Since sleep debt is a weighted sum of sleep deficits, another way to consider this effect size is by saying that one hour of sleep loss the night before was associated with 1.9% fewer hires. The average salesperson made 3.8 hires per workday and collected $936 in fees per hire. Therefore, a 1.9% decrease translates to a $67 loss per day. The average athlete experienced a 2.0% drop (1.3 points) in their game grade when they lost one hour of sleep the night before. Although these performance decreases may appear small, they can accumulate over time or across multiple people on the same team. In fact, sleep debt implies that a deficit can be spread over multiple


(a) Game Grade
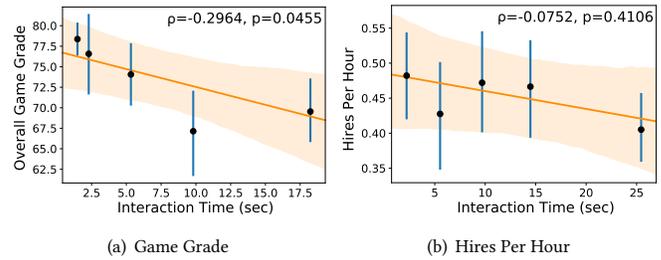
(b) Hires Per Hour

Figure 3: Regression plots showing the effect sizes between app-based performance and (a) overall game grade and (b) hires per day. App interaction time is sensitive to the job performance of the athletes, but not the salespeople.

days, so one hour of sleep loss the night before is equivalent to 2.4 hours of sleep loss a week before or 0.2 hours of sleep loss every day for a week. A more severe, but not uncommon scenario of losing an hour of sleep every day for a week is equivalent to losing 4.75 hours of sleep yesterday or 11.2 hours of sleep one week ago. On average, this loss in sleep debt was associated with a 9.5% (6.2 points) reduction in game performance, and a 9.0% ($317) reduction in hires for salespeople.

4.2 RQ.2: The Relationship Between App Interaction Time and Job Performance

Having supported the hypothesis that better sleep behavior is correlated with heightened job performance, we now explore the possibility of leveraging passively captured app interaction data as a non-invasive indicator of job performance. We investigate this question on the basis that app-based performance provides an *in-situ* measurement of psychomotor and cognitive function that may be easier to track than sleep behavior or job performance itself.

4.2.1 Analysis Procedure. To examine whether app interaction time could serve as a non-invasive indicator of job performance, we calculate the correlation between these two data sources. We also fit least squares models between app interaction time and job performance metrics to determine effect sizes. The salespeople and athletes contributed data from 122 and 46 unique days with both app interaction and job performance measurements; note that this is a many-to-one relationship since participants frequently interacted with their app multiple times in the same day.

4.2.2 Results. Figure 3 shows real-world job performance against app interaction time for those participants. App interaction time was not found to be significantly correlated with the number of hires the salespeople made (ρ=-0.0752, p=0.411). A significant correlation was found between app interaction time and the athletes' game grade (ρ=-0.296, p=0.0455). The effect size shows that athletes who were 10 seconds faster in their app interaction time had an average of 5 more points in game grades. Our app interaction metric is partly related to reaction time, so the discrepancy between athletes and salespeople in this analysis may be because the athletes' day-to-day activities require rapid, precise reactions; the salespeople's activities, on the other hand, are typically more forgiving with respect to psychomotor function. Another explanation could be that

	Raw Sleep Data			Per Person Z-Normalization of Sleep Data		
	Time-in-Bed	Sleep History	Sleep Debt	Time-in-Bed	Sleep History	Sleep Debt
Interaction Time	**-0.015** **(p=0.049)**	**-0.055** **(p=3.9×10⁻¹¹)**	**-0.095** **(p=5.2×10⁻³⁰)**	0.006 (p=0.483)	-0.012 (p=0.140)	-0.010 (p=0.230)

Table 4: Spearman correlation coefficients between sleep behavior and app-based performance. P-values are provided in parentheses; results with p-value < 0.05 are shown in bold.

STUDY 2 Statistics	Participants
Number of participants	274
Total nights of sleep tracked with app-based performance measurements	7,195
Total nights of sleep tracked	30,618
Total number of transitions between screens	16,336
Total number of times app was opened	11,140
Nights of sleep tracked per user (avg ± std)	109.2 ± 91.81
Time-in-bed in hours (avg ± std)	7.338 ± 1.628
Days of app use per user (avg ± std)	43.68 ± 46.48
App interaction time in seconds (avg ± std)	10.14 ± 10.02

Table 5: Summary statistics for our dataset in STUDY 2 after the filtering described in Section 3.6.

PFF includes contextual information, such as whether the opponent presented a favorable matchup during a game; the number of hires a salesperson can make in a given day is more dependent upon external factors (e.g., customer needs, health of the economy).

5 STUDY 2: GENERAL POPULATION

The previous study revealed statistically significant correlations between our app-based performance measurement and athletic job performance, but not salesperson job performance. These findings highlight the fact that jobs can be extremely diverse (e.g., unique skill requirements and methods of rating performance), making it challenging to test the generalizability of our findings even further. If app interaction time is truly an indicator of performance, it should be sensitive to factors that are known to impact psychomotor function. Therefore, we conducted an exploration on a broader population of 274 participants to support the idea that app-based performance truly captures aspects of a person's psychomotor and cognitive function. Using the PVT, sleep researchers have demonstrated that psychomotor and cognitive function improve with better sleep behavior [75, 76]. Separately, computing researchers have shown that the timing between interaction events in a desktop or smartphone can be an indicator of psychomotor and cognitive function [7, 87]. Our third and final research question (**RQ.3**) aims to join these two bodies of literature. Table 5 shows the summary statistics of the dataset used for this analysis after post-processing.

5.1 RQ.3: The Relationship Between App Interaction Time and Sleep Behavior

5.1.1 Analysis Procedure. It is well established that psychomotor and cognitive function vary throughout the day due to circadian rhythms homeostatic sleep drive, and sleep inertia, collectively
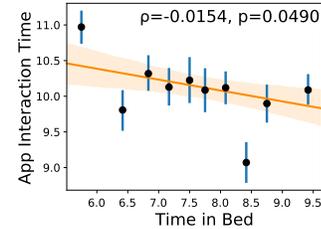
(a) Time-in-bed
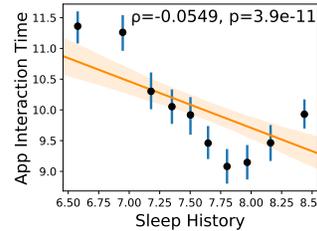
(b) Sleep History
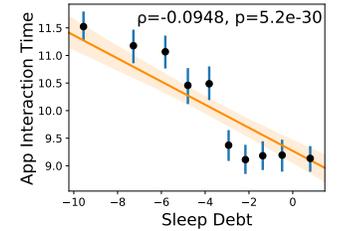
(c) Sleep Debt

Figure 4: Regression plots showing the effect sizes for the statistically significant results from Table 4. App interaction time is sensitive to many sleep metrics: (a) time-in-bed, (b) sleep history, and (c) sleep debt.

forming the three-process model of sleep [3, 7, 34, 62]. Any performance indicator should therefore be sensitive to variations of time and sleep. To examine whether this is the case for our app-based performance metric, we evaluate the relationship between app interaction time and four different measures: time of day, time since wake-up, sleep debt, and sleep history. Beyond calculating the correlation between these two data sources, we also create a generalized additive model similar to the one proposed by Althoff et al. [7] to characterize app interaction time as a function of sleep behavior and time of day. We extend this model by incorporating random effects intercepts for each user, which not only accommodates user-specific performance baselines, but also accounts for device-specific effects like the rendering capabilities of the user's smartphone. If some participants took certain medications, regularly napped, or consistently consumed high quantities of caffeine, these confounds would be adjusted through the random intercepts as well. Our participants logged 7,195 nights of sleep that were paired with at least one app interaction event during the same day.

5.1.2 Results. Table 4 summarizes the correlation coefficients between sleep behavior and app interaction time. Time-in-bed ($\rho = -0.0154\ p = 0.049$), sleep history ($\rho = -0.0549, p = 3.9 \times 10^{-11}$), and

sleep debt ($\rho = -0.0948$, $p = 5.2 \times 10^{-30}$) had negative correlations with app interaction time; in other words, participants with better sleep behaviors had faster app interaction times. Although the correlation coefficients on individual performance are rather small due to significant variation within and across participants, these estimates align with findings from previous work [7, 56, 85]. When averaging the app interaction times of samples within a certain sleep metric bin, the effect sizes are practically meaningful and span differences of up to 2.5 seconds. For example, as shown in Figure 4(c), one hour less of sleep debt was associated with app interactions times that were 0.175 seconds slower than the average. Another way to frame this effect size is that one hour less of daily sleep over the past week was associated with app interaction times that were 0.72 seconds slower. As before, the cumulative sleep behavior metrics exhibited stronger correlations than total time-in-bed; however, app-based performance correlated better with non-normalized sleep behavior metrics. This result suggests that the minimal complexity of the app interaction task engendered less variance across individuals. Moreover, we found that extended sleep does not improve psychomotor performance. Figure 4(b) shows that app interaction time was fastest when individuals had an average of 7.75 hours of daily sleep over the past week. Similar U-shaped relationships have also been reported in previous work on psychomotor performance [7] and other outcomes (e.g., mortality [53]).

Since we found statistically significant correlations between cumulative sleep behavior metrics and job performance, we used sleep history and sleep debt in generalized additive models. Figure 5 shows the variation of app interaction time as a function of time of day, time since wake-up, and the aforementioned metrics. We find that app interaction times are slowest at night and fastest between 3-6 PM; the difference between those extremes is approximately 1.5 seconds. Note that the relationship between app interaction time and time of day (Figure 5, top) generally aligns with circadian rhythm processes as measured through controlled sleep studies [3, 16, 25]. Our results also align with the chronobiological process of sleep inertia [3] since participants had slower app interaction times within one hour of waking up (Figure 5, middle). App interaction time decreases in the first six hours after wake-up and then begins to increase again, consistent with both the chronobiological process of homeostatic sleep drive [16] and previous work examining click speeds in search engines [7]. App interaction time increased by an average of 0.4 seconds when sleep history improved from 6 to 8 hours, and app interaction time increased by 0.5 seconds beyond the threshold of -5 sleep debt hours. In other words, when participants lost one hour of sleep daily for a week, they exhibited app interaction times that were 5% (0.5 seconds) slower. Note that this estimate is slightly less than the estimate of 0.72 seconds in Figure 4(c). The difference between the two estimates is explained by the fact that the generalized additive model controls for the impacts of circadian rhythm, homeostatic sleep drive, and sleep inertia, as well as participant-specific baselines through random effects.

6 DISCUSSION

Establishing the relationship between sleep behavior and job performance has been a challenge in the past due to the difficulty

(a) Accounting for Sleep History (b) Accounting for Sleep Debt
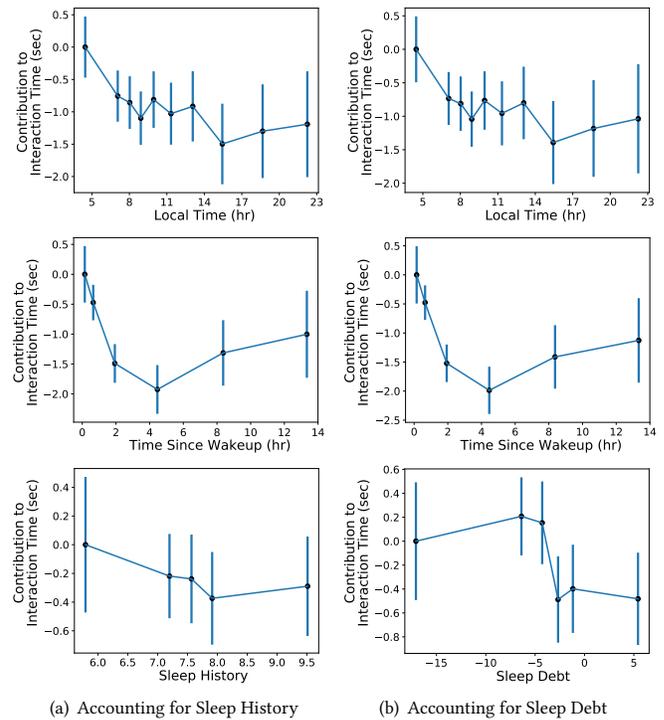
Figure 5: Generalized additive models of app interaction time accounting for (a) sleep history and (b) sleep debt. In both cases, the models account for (top row) the local time in the participant's time zone, (center row) time since wake-up, and (bottom row) sleep behavior. These models show that app interaction time is sensitive to sleep behaviors including circadian rhythm, time awake, and cumulative sleep metrics. Both models include random intercepts for each participant, and standard errors are shown.

in collecting objective measures in real-world settings. By taking advantage of ubiquitous sleep-tracking technology and the increasing desire within companies to evaluate job performance through data, our research signifies a major step towards understanding this relationship. We demonstrate that an app-based performance metric is correlated with both job performance metrics and sleep behaviors in a way that is consistent with sleep biology. This highlights an interesting opportunity for future assessments of sleep and performance in uncontrolled settings. Below, we describe the implications and limitations of our work.

6.1 Opportunities for Passive Sensing

The PVT has been used to measure psychomotor and cognitive function in the wild [2]; however, the PVT can be disruptive if deployed at inopportune moments. Other prior work has required participants to adhere to a strict sleep schedule in order to measure the effects of sleep on behavior [56, 85]. In our work, we found that our instantiation of app-based performance was correlated with both better sleep behavior and athletic job performance, suggesting the potential power of a passive, nonintrusive performance indicator. Passive sensing through ubiquitous technologies like smartphones

can enables continuous data collection for the study of populations that have traditionally been difficult to recruit to controlled studies.

We restricted our correlation analysis of app-based performance to comparable interactions within the sleep-tracking app that started from the home screen and involved single touches; however, not all interactions are created equal, nor does app interaction time tell the whole story about how the user is engaging with the app's content. Some screens require more time to process than others, and longer processing times may indicate that the user is engaging more with the displayed information. Understanding how app interaction time is a function of on-screen content could be explored further to enable more robust measurements. Beyond app interaction time, comparable performance metrics have also been elicited through other interactions like typing and web browsing [7, 82, 87]. Responses to alarms and notifications could also provide more natural opportunities for capturing app-based performance in the future.

6.2 Implications for Sleep-Tracking Apps

One design recommendation that we propose for sleep-tracking apps involves personalized views of sleep metrics. Many researchers have noted that sleep behaviors are unique according to genetic predisposition and chronotyping [4, 81]. Throughout our analyses, there were cases when normalizing sleep behavior metrics according to each user's history produced statistically significant correlations, but the same was not true for the raw data. Presenting raw values in combination with data that is scaled relative to the individual could provide useful insights to users in the future. Because sleep quality is subjective and not well-defined [39, 79], future apps could also allow users to explore what sleep metrics matter to their perceived sleep quality. In fact, we posit that job performance may be influenced by a person's perception of their own sleep quality, so our research may inform ways of exploring this matter in the future.

Finally, lapses in sleep tracking and the resulting lack of data are an important consequence of real-world data collection that should be addressed. Our dataset exhibited an extreme case of this issue since athletes can be away from home for at least 3-4 days at a time; nevertheless, travel is a regular occurrence for many people. The cumulative sleep metrics in our dataset—sleep history and sleep debt—were most informative in our analyses related to sleep behavior. We used the average time-in-bed of nearby nights for imputation when a participant skipped a night of sleep tracking (Section 3.6). Future work could explore other alternatives to imputation, such as improving generative models through deep learning [30, 94] or multi-device sensing to remedy data gaps [50].

6.3 Additional Context Information

PFF game grades are able to incorporate context because they are assigned by experts who watch the games and understand the athletes' match. Our other data streams, however, lacked such context. For example, the performance of salespeople depends on the demand of their goods and services. Job performance in general is also a function of experience and division of labor. Such information from managers and worker profiles could be incorporated for refined analyses in future work.

Sleep is known to be affected by a wide variety of factors: age [24, 31, 93], ambient light [52], caffeine intake [54], and diet [38], to name a few. The effect of travel between time zones (2–3 hour difference) has not been shown to significantly impact sleep [80], but an effect has been demonstrated on athletic performance [43]. Measuring these factors through sensors and accounting for their effects in statistical analyses could improve evidence of links between sleep behavior, job performance, and app usage.

6.4 Limitations

Our dataset included participants from a bankruptcy law firm consultancy and the NFL, which allowed us to compare two populations with distinct job demands whose job performance can be quantified effectively. In both cases, we were able to identify sleep behavior metrics that correlated with job performance; however, the correlations manifested in different sleep behavior metrics (e.g., sleep debt for salespeople, personalized sleep history for athletes) (Section 4.1. Beyond the discrepancy between the two groups' job demands, the differences in results can also be attributed to idiosyncrasies within the job performance metrics themselves. For the salespeople, the number of hires an employee is able to make may depend on the state of the economy and the rate of bankruptcy in the country. For the athletes, the subjective nature of the expert's grades can manifest in anchoring effects towards common values [84]. We use rank-based correlation methods and per-person normalization to account for some of these idiosyncrasies (Section 4.1), but future work should explore and compare alternative sources of job performance data. Furthermore, an exciting avenue of research may entail the creation of a job performance metric that generalizes across different careers.

Although salespeople and athletes have very different job demands, they do not cover the entire spectrum of careers. Each profession has its own demands and may not overlap with either of the ones that were included in our study. There was also an element of selection bias in our participant pool; the people who enrolled in our observational study may have been more excited to track their sleep and interact with the app than the average person, producing inflated app engagement measurements. Similarly, the observational and correlational nature of our data preclude us from making causal inferences. Learning about how our findings may generalize to other populations remains an area of future work.

Lastly, there are many confounds that could have affected our datasets. People have unique habits that affect their sleep behavior and job performance [38, 52, 54]. Unique smartphone parameters like clock speed or operating system throttling due to current battery level affect app interaction time. We addressed within-person confounds as much as possible via statistical methods. For our correlational analyses, we examined both raw and per-person normalized sleep behavior metrics (Section 4.1, 4.2, 5.1). For our generalized additive model of app interaction time against sleep behavior and time of day, we utilized random effects intercepts to accommodate for performance baselines, habits, and device specifications specific to each participant (Section 5.1). These steps helped us account for confounds that existed throughout a participant's enrollment in the study, including regular medication intake, naps, or caffeine consumption.

7 CONCLUSION

Many people recognize that improving sleep behavior benefits job performance, but the precise relationship between the two has been difficult to capture and quantify in the past. Our study advances the literature in this space by providing a correlational analysis between objectively measured sleep behavior metrics from a mattress sensor and job performance metrics from a bankruptcy law firm and the NFL. Our findings suggest that establishing good sleep behaviors over extended periods is more important to job performance than simply getting a good night's sleep one day prior. We also found evidence that passively captured app interaction metrics can serve as a useful indicator for some job performance and sleep measures, thereby highlighting another mechanism through which researchers can collect relevant psychomotor and cognitive performance measures at scale. It is our hope that our work inspires researchers to examine *in-situ* sleep behaviors and performance measures across diverse contexts to further develop our understanding of human performance.

ACKNOWLEDGMENTS

REFERENCES

[1] Saeed Abdullah, Mark Matthews, Elizabeth L. Murnane, Geri Gay, and Tanzeem Choudhury. 2014. Towards circadian computing: "Early to bed and early to rise" makes some of us unhealthy and sleep deprived. In *Proc. UbiComp '14*. 673–684.

[2] Saeed Abdullah, Elizabeth L Murnane, Mark Matthews, Matthew Kay, Julie A Kientz, Geri Gay, and Tanzeem Choudhury. 2016. Cognitive rhythms: unobtrusive and continuous sensing of alertness using a mobile phone. In *Proc. UbiComp '16*. ACM Press, New York, New York, USA, 178–189.

[3] Torbjörn Åkerstedt and Simon Folkard. 1997. The three-process model of alertness and its extension to performance, sleep latency, and sleep length. *Chronobiology International* 14, 2 (jan 1997), 115–123.

[4] Karla V Allebrandt et al. 2010. CLOCK Gene Variants Associate with Sleep Duration in Two Independent Populations. *Biological Psychiatry* 67, 11 (jun 2010).

[5] Tim Althoff. 2017. Population-scale pervasive health. *IEEE Pervasive Computing* 16, 4 (2017).

[6] Tim Althoff, Eric Horvitz, and Ryen W White. 2018. Psychomotor function measured via online activity predicts motor vehicle fatality risk. *npj Digital Medicine* 1, 1 (2018).

[7] Tim Althoff, Eric Horvitz, Ryen W White, and Jamie Zeitzer. 2017. Harnessing the Web for Population-Scale Physiological Sensing. In *Proc. WWW '17*. 113–122.

[8] Sonia Ancoli-Israel, Roger Cole, Cathy Alessi, Mark Chambers, William Moorcroft, and Charles P Pollak. 2003. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* 26, 3 (2003), 342–392.

[9] Consumer Electronics Association and National Sleep Foundation. 2015. *Consumer Awareness and Perception of Sleep Technology*. Consumer Electronics Association.

[10] Sangwon Bae, Denzil Ferreira, Brian Suffoletto, Juan C Puyana, Ryan Kurtz, Tammy Chung, and Anind K Dey. 2017. Detecting Drinking Episodes in Young Adults Using Smartphone-based Sensors. *Proc. IMWUT '17* 1, 2 (jun 2017), 1–36.

[11] Joseph Baranski and Ross Pigeau. 1997. Self-monitoring cognitive performance during sleep deprivation: effects of modafinil, d-amphetamine and placebo. *Journal of Sleep Research* 6, 2 (jun 1997), 84–91.

[12] Mathias Basner, Kenneth M Fomberstein, Farid M Razavi, Siobhan Banks, Jeffrey H William, Roger R Rosa, and David F Dinges. 2007. American time use survey: sleep time and its relationship to waking activities. *Sleep* 30, 9 (2007).

[13] Jared Bauer, Sunny Consolvo, Benjamin Greenstein, Jonathan Schooler, Eric Wu, Nathaniel F Watson, and Julie Kientz. 2012. ShutEye: Encouraging Awareness of Healthy Sleep Recommendations with a Mobile, Peripheral Display. In *Proc. CHI '12*. ACM Press, New York, New York, USA, 1401.

[14] Gregory Belenky et al. 2003. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of Sleep Research* 12, 1 (mar 2003), 1–12.

[15] Mark Blagrove, Carol Alexander, and James A Horne. 1995. The effects of chronic sleep reduction on the performance of cognitive tasks sensitive to sleep deprivation. *Applied Cognitive Psychology* 9, 1 (feb 1995), 21–40.

[16] Alexander A Borbély, Serge Daan, Anna Wirz-Justice, and Tom Deboer. 2016. The two-process model of sleep regulation: A reappraisal. *Journal of Sleep Research* 25, 2 (2016), 131–143.

[17] David H Brendel et al. 1990. Sleep Stage Physiology, Mood, and Vigilance Responses to Total Sleep Deprivation in Healthy 80-Year-Olds and 20-Year-Olds. *Psychophysiology* 27, 6 (nov 1990), 677–685.

[18] Nikhil Byanna and Diego Klabjan. 2016. Evaluating the Performance of Offensive Linemen in the NFL. *arXiv preprint arXiv:1603.07593* (2016).

[19] Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tanzeem Choudhury, and Andrew T Campbell. 2013. Unobtrusive sleep monitoring using smartphones. In *Proc. PervasiveHealth '13*. IEEE, 145–152.

[20] Ronald D Chervin. 2000. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea. *Chest* 118, 2 (2000), 372–379.

[21] Ralph B D'Agostino. 1971. An omnibus test of normality for moderate and large size samples. *Biometrika* 58, 2 (1971), 341–348.

[22] Nediyana Daskalova, Bongshin Lee, Jeff Huang, Chester Ni, and Jessica Lundin. 2018. Investigating the Effectiveness of Cohort-Based Sleep Recommendations. *Proc. IMWUT '18* 2, 3 (2018), 1–19.

[23] Nediyana Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, Nicole Nugent, Julie Boergers, John McGeary, and Jeff Huang. 2016. SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations. In *Proc. UIST '16*. 347–358.

[24] Derk Jan Dijk and Jeanne F Duffy. 1999. Circadian regulation of human sleep and age-related changes in its timing, consolidation and EEG characteristics.

[25] Derk Jan Dijk, Jeanne F Duffy, and Charles A Czeisler. 1992. Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *Journal of Sleep Research* 1, 2 (1992), 112–117.

[26] David F Dinges. 2004. Sleep debt and scientific evidence. *Sleep* 27, 6 (sep 2004).

[27] David F Dinges, Frances Pack, Katherine Williams, Kelly A Gillen, John H Powell, Goeffrey E Ott, Caitlin Aptowicz, and Allen I Pack. 1997. *Cumulative Sleepiness, Mood Disturbance, and Psychomotor Vigilance Performance Decrements During a Week of Sleep Restricted to 4-5 Hours per Night*. Technical Report 4.

[28] David F Dinges and John W Powell. 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers* 17, 6 (1985), 652–655.

[29] Christopher C Dodson, Eric S Secrist, Suneel B Bhat, Daniel P Woods, and Peter F Deluca. 2016. Anterior Cruciate Ligament Injuries in National Football League Athletes From 2010 to 2013: A Descriptive Epidemiology Study. *Orthopaedic Journal of Sports Medicine* 4, 3 (mar 2016).

[30] Chenguang Fang and Chen Wang. 2020. Time Series Data Imputation: A Survey on Deep Learning Approaches. arXiv:cs.LG/2011.11347

[31] Irwin Feinberg. 1974. Changes in sleep cycle patterns with age. *Journal of Psychiatric Research* 10, 3-4 (1974), 283–306.

[32] Raihana Ferdous, Venet Osmani, and Oscar Mayora. 2015. Smartphone app usage as a predictor of perceived stress levels at workplace. In *Proc. PervasiveHealth 2015*. Institute of Electrical and Electronics Engineers Inc., 225–228. arXiv:1803.03863

[33] Michael Gertz. 2017. NFL Census 2016 - ProFootballLogic.

[34] Namni Goel, Mathias Basner, Hengyi Rao, and David F Dinges. 2013. Circadian rhythms, sleep deprivation, and human performance. In *Progress in Molecular Biology and Translational Science*. Elsevier, 155–190.

[35] Scott A Golder and Michael W Macy. 2011. Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science* 333, 6051 (2011), 1878–1881.

[36] Mitchell L Gordon, Leon Gatys, Carlos Guestrin, Jeffrey P Bigham, Andrew Trister, and Kayur Patel. 2019. App Usage Predicts Cognitive Ability in Older Adults. In *Proc. CHI '19*. 168.

[37] G. Guerrero-Mora, Palacios Elvia, A. M. Bianchi, J. Kortelainen, M. Tenhunen, S. L. Himanen, M. O. Mendez, E. Arce-Santana, and O. Gutierrez-Navarro. 2012. Sleep-wake detection based on respiratory signal acquired through a Pressure Bed Sensor. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. 3452–3455.

[38] Shona L Halson. 2008. Nutrition, sleep and recovery. *European Journal of Sport Science* 8, 2 (mar 2008), 119–126.

[39] Allison G Harvey, Kathleen Stinson, Katriina L Whitaker, Damian Moskovitz, and Harvinder Virk. 2008. The Subjective Meaning of Sleep Quality: A Comparison of Individuals with and without Insomnia. *Sleep* 31, 3 (mar 2008), 383–393.

[40] Jennifer L Hicks, Tim Althoff, Peter Kuhar, Bojan Bostjancic, Abby C King, Jure Leskovec, Scott L Delp, et al. 2019. Best practices for analyzing large-scale health data from wearables and smartphone apps. *NPJ digital medicine* 2, 1 (2019).

[41] Jim Horne. 2004. Is there a sleep debt? *Sleep* 27, 6 (sep 2004), 1047–1049.

[42] Vanessa Ibáñez, Josep Silva, and Omar Cauli. 2018. A survey on sleep assessment methods. *PeerJ* 6 (2018), e4849.

[43] Richard Jehue, David Street, and Robert Huizenga. 1993. Effect of time zone and game time changes on team performance: National Football League. *Medicine and Science in Sports and Exercise* 25, 1 (jan 1993), 127–131.

[44] Megan E Jewett, Derk Jan Dijk, Richard E Kronauer, and David F Dinges. 1999. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep* 22, 2 (1999), 171–179.

[45] William DS Killgore, Thomas J Balkin, and Nancy J Wesensten. 2006. Impaired decision making following 49 h of sleep deprivation. *Journal of Sleep Research* 15, 1 (mar 2006), 7–13.

[46] William DS Killgore, Ellen T Kahn-Greene, Erica L Lipizzi, Rachel A Newman, Gary H Kamimori, and Thomas J Balkin. 2008. Sleep deprivation reduces perceived emotional intelligence and constructive thinking skills. *Sleep Medicine* 9, 5 (jul 2008), 517–526.

[47] Shingo Kitamura et al. 2016. Estimating individual optimal sleep duration and potential sleep debt. *Scientific Reports* 6, 1 (dec 2016), 35812.

[48] Elizabeth B Klerman and Derk Jan Dijk. 2008. Age-Related Reduction in the Maximal Capacity for Sleep-Implications for Insomnia. *Current Biology* 18, 15 (2008), 1118–1123.

[49] Kristen L Knutson, Eve Van Cauter, Paul J Rathouz, Thomas DeLeire, and Diane S Lauderdale. 2010. Trends in the prevalence of short sleepers in the USA: 1975-2006. *Sleep* 33, 1 (2010), 37–45.

[50] Ping-Ru T Ko, Julie A Kientz, Eun Kyoung Choe, Matthew Kay, Carol A Landis, and Nathaniel F Watson. 2015. Consumer Sleep Technologies: A Review of the Landscape. *Journal of Clinical Sleep Medicine* 11, 12 (2015), 1455–1461.

[51] Juha M. Kortelainen, Martin O. Mendez, Anna Maria Bianchi, Matteo Matteucci, and Sergio Cerutti. 2010. Sleep staging based on signals acquired through bed sensor. *IEEE Transactions on Information Technology in Biomedicine* 14, 3 (may 2010), 776–785.

[52] Tomoaki Kozaki, Shingo Kitamura, Yuichi Higashihara, Keita Ishibashi, Hiroki Noguchi, and Akira Yasukouchi. 2005. Effect of Color Temperature of Light Sources on Slow-wave Sleep. *Journal of Physiological Anthropology and Applied Human Science* 24, 2 (2005), 183–186.

[53] Daniel F Kripke, Ruth N Simons, Lawrence Garfinkel, and E Cuyler Hammond. 1979. Short and long sleep and sleeping pills: is increased mortality associated? *Archives of general psychiatry* 36, 1 (1979), 103–116.

[54] Hans Peter Landolt, Derk-Jan Dijk, Stephanie E Gaus, and Alexander A Borbély. 1995. Caffeine reduces low-frequency delta activity in the human sleep EEG. *Neuropsychopharmacology* 12, 3 (1995), 229–238.

[55] F. C. Leone, L. S. Nelson, and R. B. Nottingham. 1961. The Folded Normal Distribution. *Technometrics* 3, 4 (1961), 543–550.

[56] June C Lo, Ju Lynn Ong, Ruth LF Leong, Joshua J Gooley, and Michael WL Chee. 2016. Cognitive Performance, Sleepiness, and Mood in Partially Sleep Deprived Adolescents: The Need for Sleep Study. *Sleep* 39, 3 (2016), 687–698.

[57] Cheri D Mah, Kenneth E Mah, Eric J Kezirian, and William C Dement. 2011. The Effects of Sleep Extension on the Athletic Performance of Collegiate Basketball Players. *Sleep* 34, 7 (jun 2011), 943–950.

[58] Alex Mariakakis, Sayna Parsi, Shwetak N Patel, and Jacob O Wobbrock. 2018. Drunk User Interfaces: Determining Blood Alcohol Level through Everyday Smartphone Tasks. In *Proc. CHI '18*, Vol. l. 1–13.

[59] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, and Paul Johns. 2014. Bored mondays and focused afternoons: The rhythm of attention and online activity in the workplace. In *Proc. CHI '14*. 3025–3034.

[60] Bruce J Martin. 1981. Effect of sleep deprivation on tolerance of prolonged exercise. *European Journal of Applied Physiology and Occupational Physiology* 47, 4 (dec 1981), 345–354.

[61] Bruce J Martin and Gary M Gaddis. 1981. Exercise after sleep deprivation. *Medicine and Science in Sports and Exercise* 13, 4 (1981), 220–223.

[62] Robert L Matchock and J Toby Mordkoff. 2009. Chronotype and time-of-day influences on the alerting, orienting, and executive components of attention. *Experimental brain research* 192, 2 (jan 2009), 189–198.

[63] Giuliana Mazzoni, S Gori, G Formicola, C Gneri, R Massetani, L Murri, and P Salzarulo. 1999. Word recall correlates with sleep cycles in elderly subjects. *Journal of Sleep Research* 8, 3 (sep 1999), 185–188.

[64] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. Towards multimodal anticipatory monitoring of depressive states through the analysis of human-smartphone interaction. In *Proc. UbiComp '16*. ACM, 1132–1138.

[65] Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I. Hong. 2014. Toss 'n' turn: smartphone as sleep and sleep quality detector. In *Proc. CHI '14*. 477–486.

[66] Elizabeth L Murnane et al. 2016. Mobile manifestations of alertness. In *Proc. MobileHCI '16*. ACM Press, New York, USA.

[67] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing* 16, 1 (2012).

[68] GF Pickett and AF Morris. 1975. Effects of acute sleep and food deprivation on total body response time and cardiovascular performance. *The journal of sports medicine and physical fitness* 15, 1 (mar 1975), 49–56.

[69] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When attention is not scarce-detecting boredom from mobile phone usage. In *Proc. UbiComp '15*. 825–836.

[70] Emma Pierson, Tim Althoff, Daniel Thomas, Paula Hillard, and Jure Leskovec. 2021. Daily, weekly, seasonal and menstrual cycles in women's mood, behaviour and vital signs. *Nature Human Behaviour* (2021).

[71] June J Pilcher and Allen I Huffcutt. 1996. Effects of sleep deprivation on performance. *Sleep* 19, 4 (1996), 318–326.

[72] Pro Football Focus. 2017. How We Grade. , 4 pages.

[73] Matthew T Provencher et al. 2018. A History of Anterior Cruciate Ligament Reconstruction at the National Football League Combine Results in Inferior Early National Football League Career Participation. *Arthroscopy - Journal of Arthroscopic and Related Surgery* 34, 8 (2018), 2446–2453.

[74] Tauhidur Rahman et al. 2015. DoppleSleep: A contactless unobtrusive sleep sensing system using short-range doppler radar. In *Proc. UbiComp '15*. 39–50.

[75] Pooja Rajdev, David Thorsley, Srinivasan Rajaraman, Tracy L Rupp, Nancy J Wesensten, Thomas J Balkin, and Jaques Riefman. 2013. A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation. *Journal of theoretical biology* 331 (2013), 66–77.

[76] Sridhar Ramakrishnan, Srinivas Laxminarayan, David Thorsley, Nancy J Wesensten, Thomas J Balkin, and Jaques Reifman. 2012. Individualized performance prediction during total sleep deprivation: Accounting for trait vulnerability to sleep loss. In *Proc. EMBS '12*. 5574–5577.

[77] Sridhar Ramakrishnan, Nancy J Wesensten, Thomas J Balkin, and Jaques Reifman. 2016. A Unified Model of Performance: Validation of its Predictions across Different Sleep/Wake Schedules. *Sleep* 39, 1 (2016), 249–262.

[78] Jukka Ranta, Timo Aittokoski, Mirja Tenhunen, and Mikko Alasaukko-Oja. 2019. EMFIT QS heart rate and respiration rate validation. *Biomedical Physics and Engineering Express* 5, 2 (2019), 25016.

[79] Ruth Ravichandran, Sang Wha Sien, Shwetak N Patel, Julie A Kientz, and Laura R Pina. 2017. Making sense of sleep sensors: How sleep sensing technologies support and undermine sleep health. In *Proc. CHI '17*. ACM, 6864–6875.

[80] Louise K Richmond, Brian Dawson, Glenn Stewart, Stuart Cormack, David R Hillman, and Peter R Eastwood. 2007. The effect of interstate travel on the sleep patterns and performance of elite Australian Rules footballers. *Journal of Science and Medicine in Sport* 10, 4 (jun 2007), 252–258.

[81] Till Roenneberg, Anna Wirz-Justice, and Martha Merrow. 2003. Life between clocks: Daily temporal patterns of human chronotypes. *Journal of Biological Rhythms* 18, 1 (feb 2003), 80–90.

[82] Martin Thirkettle, Jennifer Lewis, Darren Langdridge, and Graham Pike. 2018. A Mobile App Delivering a Gamified Battery of Cognitive Tests Designed for Repeated Play (OU Brainwave): App Design and Cohort Study. *JMIR Serious Games* 6, 4 (2018), e10519.

[83] Eirunn Thun, Bjørn Bjorvatn, Elisabeth Flo, Anette Harris, and Ståle Pallesen. 2015. Sleep, circadian rhythms, and athletic performance. *Sleep Medicine Reviews* 23 (oct 2015), 1–9.

[84] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 4157 (1974), 1124–1131.

[85] Hans PA Van Dongen, Naomi L Rogers, and David F Dinges. 2003. Sleep debt: Theoretical and empirical issues.

[86] T Van Helder and Marek W Radoki. 1989. Sleep Deprivation and the Effect on Exercise Performance. *Sports Medicine* 7, 4 (apr 1989), 235–247.

[87] Lisa M Vizer, Lina Zhou, and Andrew Sears. 2009. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human Computer Studies* 67, 10 (oct 2009), 870–886.

[88] Olivia J Walch, Amy Cochran, and Daniel B Forger. 2016. A global quantification of "normal" sleep schedules using smartphone data. *Science advances* 2, 5 (2016).

[89] Matthew P Walker and Robert Stickgold. 2005. Sleep, Memory, and Plasticity. *Annual Review of Psychology* 57, 1 (jan 2005), 139–166.

[90] Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T Campbell. 2015. SmartGPA: How smartphones can assess and predict academic performance of college students. In *Proc. UbiComp '15*. 1–13.

[91] Andrew M Watson. 2017. Sleep and Athletic Performance. *Current Sports Medicine Reports* 16, 6 (2017), 413–418.

[92] Ann M Williamson and Anne-Marie Feyer. 2000. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational and Environmental Medicine* 57, 10 (2000), 649–655.

[93] In Young Yoon, Daniel F Kripke, Jeffrey A Elliott, Shawn D Youngstedt, Katharine M Rex, and Richard L Hauger. 2003. Age-related changes of circadian rhythms and sleep-wake cycles. *Journal of the American Geriatrics Society* 51, 8 (aug 2003), 1085–1091.

[94] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2018. GAIN: Missing Data Imputation using Generative Adversarial Nets. arXiv:cs.LG/1806.02920