# A Multi-sensor Approach to Linking Behavior to Job Performance

Alison M. Perez[(✉)], Amanda E. Kraft, Raquel Galvan-Garza,
Matthew Pava, Amanda Barkan, William D. Casebeer,
and Matthias D. Ziegler

Advanced Technology Laboratories, Lockheed Martin,
Arlington, VA 22203, USA
Alison.M.Perez@lmco.com

**Abstract.** Traditionally job performance reviews occur infrequently, only a few times a year at best, and can be largely subjective. Additionally, most quantitative assessment of job performance (e.g., hours at work, number of articles published, etc.) do not give a complete picture, either because they do not account for individual differences or job variability, or they rely only on single measures, subjective reporting, sparse performance measurements or a combination of these factors. Here we report on our initial comparison of objective signals obtained from unobtrusive physiologic and environmental sensors to self-reports of workplace performance and wellbeing. Our results provide evidence that objective metrics of physiological and environmental factors for individuals might be useful in supplementing subjective reports of workplace performance and wellbeing. We posit that a large longitudinal study would provide enough information to automate timely analysis that would allow for tailored performance interventions, workforce retention, and mitigation of negative workplace behaviors.

**Keywords:** Job performance · Workplace behavior · Physiological monitoring
Wellbeing · Environmental monitoring

## 1 Introduction

Job performance measurements allow both employers and employees to understand their workplace capabilities over time and can be informative in setting expectations for future workplace behavior and wellbeing. In today's workforce, employees have diverse responsibilities and often face task-saturation [1]. Creating a system that provides multimodal, objective and timely assessments of workplace performance would allow for a comprehensive solution to the challenge of caring for a highly productive but task-saturated workforce. This study is an initial step towards investigating an extensive suite of sensors and metrics in predicting objective and subjective reports of performance and wellbeing. The results of this work will be used for identifying possible connections between signals and self-report assessments as well as being an important building block for planning for and mitigating problems that may occur in larger studies. The lessons learned from this pilot experiment allows us to run a large N

study that can inform changes made to the workplace as a whole as well as in tailoring interventions to individuals that may be impacted by factors of the workplace differently.

Informative metrics of workplace performance and the factors that influence performance and wellbeing can inform interventions and lead to improvements for both employees and employers [2]. Our study compares measures from unobtrusive physiologic and environmental sensors to self-reports of workplace performance and wellbeing (See Table 1). While many studies focus on individual behavioral and physiological measures contributing to individual differences, our sensor suite also includes an often overlooked aspect of job performance by sensing environmental factors that can improve the accuracy of predictions about factors contributing to workplace performance and wellbeing. For instance, ambient $CO_2$ levels affect cardiac measures, leading to fluctuations in heart rate or HRV that are (1) unrelated to physical exertion or psychological arousal (e.g., Anxiety); and (2) may be independently related to workplace performance outcomes (e.g., reduced higher-order decision-making) [3, 4]. In the first case, adding a $CO_2$ measurement to the dataset should reduce error in the cardiac-based signals to improve the predictive value of cardiac signals on performance (i.e., convergent validity). In the second case – if $CO_2$ levels indeed affect decisions influencing workplace performance – adding a $CO_2$ measurement should also boost the predictive validity (i.e., criterion validity) of the overall performance model.

**Table 1.** Components of the sensor suite and their relative measurement variables

| Sensor | Fitbit | Zephyr | Air Quality Egg | Actiwatch |
|---|---|---|---|---|
| Signal type | Individual | Individual | Environmental | Individual, Environmental |
| Signals | Activity Heart rate Sleep quality | ECG Breathing Posture Activity | CO2 conc. Temperature Rel. Humidity | Sleep quantity and quality Natural, artificial, and total light exposure Activity |

In this pilot study we identify salient features from our sensor suite that correspond best with ground truth measures of workplace performance, health and wellbeing self-reports through measures of correlation. We will then discuss our plans to use a dual-track approach, combining theory-driven and data-driven models, to link known and unknown connections between sensor signal features to individual job performance variables. These results are intended to support a now ongoing study including 258 participants.

## 2   Methods

### 2.1   Participants

Six participants were equipped with an array of sensors, further described below, including: Actiwatch, Fitbit, Zephyr, and an Air Quality Egg over the course of

3 weeks. All participants completed Daily Ground Truth Batteries (DGTB), composed of survey questions on workplace performance, health and wellbeing. All participants were native speakers of English and had at least a high school education. Participants signed the informed consent after reading the experiment summary. The number of daily survey responses per participant ranged from 2 to 21, with an average of 12 responses per participant. The study was approved by and conducted in accordance with the standards of the Western Internal Review Board.

## 2.2 Sensors

The following suite of sensors was used for this study (For a summary of the sensors, see Table 1).

The Zephyr bioharness is a chest-worn sensor which includes an accelerometer and stretch sensors for breathing rate. It measures three-dimensional postural data and provides ECG (electrocardiogram) and associated heart-related data. While wrist-worn sensors provide some of the same measures, ECG is far more accurate and provides more fidelity than wrist-worn, optical heart rate detection methods [5]. Participants were not required, but had the option to wear the Zephyr bioharness overnight if it did not disrupt their sleep.

The Fitbit Charge 2 is a low-profile wrist-worn activity tracker that can validate many of the same signals as the Zephyr bioharness. It can be worn 24-h to provide measures such as heart rate, sleep quality and other individual measurements of out-of-office activity that may influence wellbeing, health, affect, and burnout. In a recent study comparing wrist-worn trackers to laboratory grade ECG, the Apple Watch, Mio FUSE and Fitbit Charge 2 scored the highest among all sensors tested [5].

The Air Quality Egg is a commercially available environmental sensor that sits on the employee's desk and measures $CO_2$ concentration in the air, temperature and relative humidity. $CO_2$ concentration affects HRV and cerebral blood flow increases in response to chronic low levels of $CO_2$ which in turn can impact executive function and anxiety [3, 6].

The Actiwatch Spectrum Plus is a commercial wrist-worn sensor that is a research-grade sleep quality monitor. Sleep quality (not just time in bed) correlates with physiological health complaints ($r2 = 0.39$–$0.60$) [7]. Also, Sleep and light affect hormones related to overeating behavior [8]. Additionally, there are higher rates of sleep disturbance among those with burnout ($\eta2 = .396$), and sleep deprivation can affect executive function and anxiety and contributes to deficits in job performance [9–11]. This sensor also detects the light spectrum to allow quantification of an individual's average exposure to different red, green and blue compositions of light. Not only is light exposure linked to depression irrespective of activity levels, but also it is predictive of effect and how seasons and shift-work affect workplace performance [12, 13].

## 2.3 Experimental Procedure

Participants received the Fitbit Charge 2, Air Quality Egg, the Zephyr bioharness, and Actiwatch Spectrum Plus sensors at the beginning of the study and were trained on how to use all sensors. Participants picked up the Zephyr puck at the beginning of each work

shift at a check-in desk positioned within the facility to maximize convenience and boost throughput during this process. Participants deposited the Zephyr puck prior to leaving work so researchers could upload the data and recharge overnight. Participants who opted in for wearing the device overnight were provided with a second puck when they left work. Participants wore the Actiwatch and Fitbit continuously (24 h a day, for three weeks). Participants also completed daily surveys (DGTB).

## 2.4 Metrics

The DGTB consisted of various self-report surveys covering topics such as on workplace performance, health and wellbeing. Subsets of the DGTB were administered each day to reduce participant burden. For a summary please see Table 2.

**Table 2.** Series of questionnaires included in the daily survey batteries

| **Daily Ground Truth Battery** |
| :---: |
| In-Role Behavior scale (IRB) |
| Individual Task Proficiency scale (ITP) |
| Organizational Citizenship Behavior (OCB)/Counterproductive Work Behavior Scale (CWB) |
| Big Five Inventory-10 |
| Alcohol quantity |
| Tobacco quantity |
| Physical activity durations |
| Sleep duration |
| Shortened Positive and Negative Affect Schedule Expanded Form (PANAS) |
| Anxiety rating |
| Stress rating |
| Social, Activity and Location Context |

## 2.5 Analyses

Given the limited number of participants, correlations were produced to demonstrate patterns in the data that are of interest. Correlations were derived between daily survey responses and corresponding features from Zephyr, Fitbit, Air Quality Egg, and Actiwatch. Depending on the compliance of each individual wearing each sensor and answering each daily survey, the number of data points differ across each correlation graph. The number of data points is derived the number of surveys taken when each sensor was being used over the course of the 3-week data collection. Features from each of the sensors were averaged over the course of the day of the survey response, except for health survey items, as these items asked about the prior day.

Features from Zephyr were averaged over the course of a day and were then correlated with survey responses (N = 66).

Features from Fitbit were averaged over the course of a day and correlated with a subset of the daily survey scores (N = 61): Stress, Anxiety, PANAS positive, PANAS negative, and PANAS all (net score), Alcohol, and Exercise. For Alcohol and Exercise, Fitbit data from the prior day is used. All other labels in the correlation plot correspond to Fitbit features.

Features from Air Quality Egg were averaged over 5 (N = 39), 15 (N = 41), and 30 (N = 42) minute windows preceding the submission of daily survey scores for Stress, Anxiety, and PANAS. Features were averaged over 9 a.m.–5 p.m. for Job Survey scores (N = 21), and averaged over 9 a.m.–5 p.m. the prior day for Health Survey scores (N = 30). Since the Egg only monitors while in the office, weekend data was excluded from the correlations.

Features from Actiwatch were averaged over the day and correlated with the following daily survey scores (N = 58): Stress, Anxiety, PANAS positive, PANAS negative, and PANAS all (net score). All other labels in the correlation plot correspond to Actiwatch features.

## 3   Results

A total of 66 instances of Zephyr features and DGTB responses were correlated across the 6 participants. Some correlations align with the expected direction, concurrent with previous literature, although stronger or weaker given the small number of individuals included in this pilot study [14, 15]. For example, the DGTB item "social_context" options were ordered by decreasing social activity: 1 indicated verbal interactions, 2 indicated written interactions, and 3 indicated no interactions. As expected, there is an inverse correlation with breathing rate, reflecting an increase in breathing rate with speech and reduction with reduced social interaction [14]. However, there are other correlations not expected, such as higher heart rate correlating with an increased number of hours slept. Based on preliminary reviews of this data, this relationship appears to result from a bias of one individual who consistently recorded both higher resting heart rate and the number of hours slept in comparison to the other participants. While raw heart rate was used for these correlations, normalizing heart rates against individual resting rates would likely reduce this type of bias in subsequent analyses. All participants were non-smokers, resulting in a lack of variation in tobacco DGTB responses (See Fig. 1).

A total of 61 instances of Fitbit features and DGTB responses were correlated across the 6 participants. Again, given the limited number of individuals included in the pilot study, our results may not have enough power to show subtle but informative relationships between variables. For example, the Fitbit feature of minutes awake shows a moderate positive correlation with both stress and anxiety, however, given a larger sample we would assume this correlation would strengthen. Other relationships that were expected such as sleep efficiency and exercise showed low to no relation from the correlational analysis (See Fig. 2).
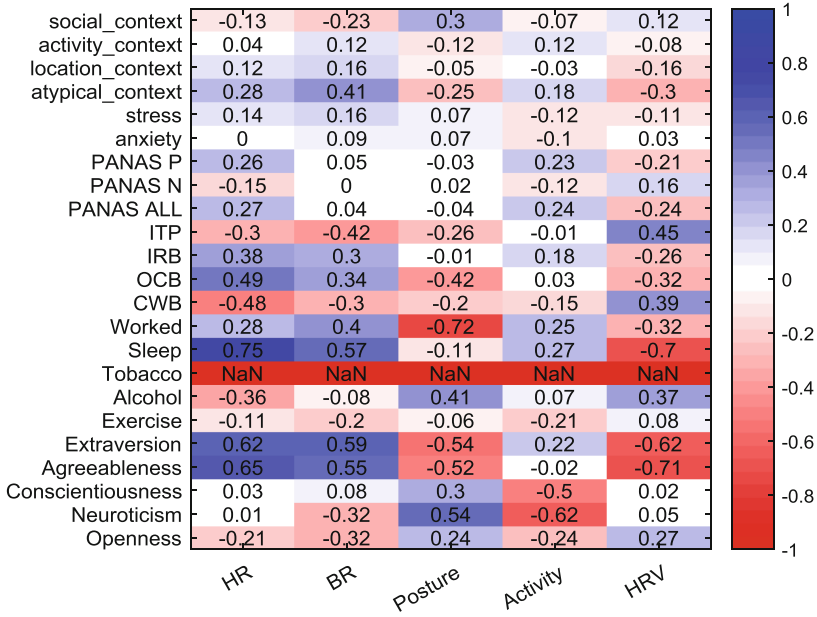
**Fig. 1.** Correlations between DGTB and Zephyr. NaNs represent that there were no tobacco users in the subject group
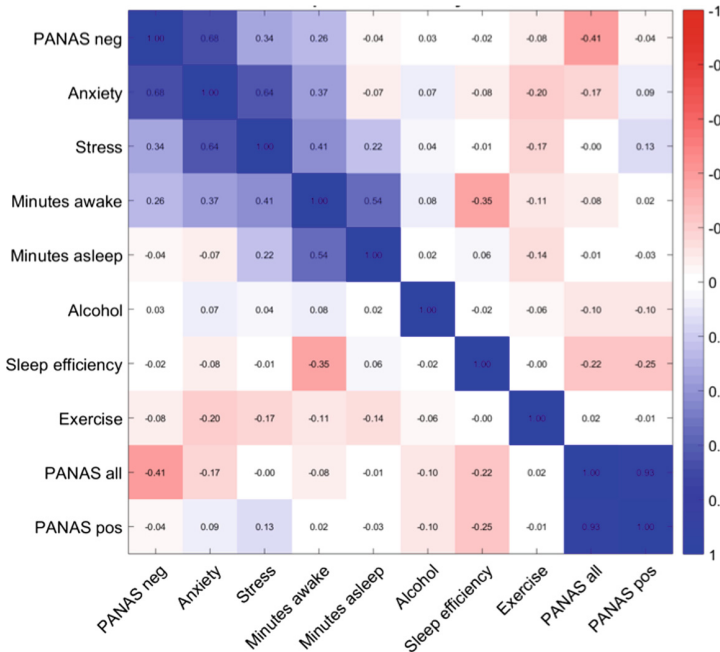


**Fig. 2.** Correlations between DGTB and Fitbit daily averages

Since the Air Quality Egg only monitors while in the office, weekend data was excluded from the correlations. Labels on the x-axes correspond to daily survey scores, while labels on the y-axes correspond to Egg features. Some correlations were unexpected, such as that between high temperatures and less counterproductive work performance. Some correlations do align with the expected direction, such as the modest correlation between $CO_2$ composition and higher exercise, and higher $CO_2$ and lower negative effect (See Fig. 3) [16, 17].
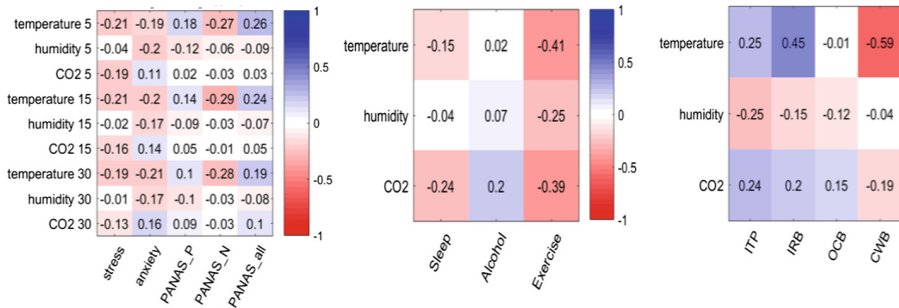


**Fig. 3.** Correlations between Air Quality Egg and Left: Daily survey items Center: Health survey items and Right: Job survey items

A total of 58 instances of Actiwatch features and DGTB responses were correlated across the 6 participants. High sleep fragmentation, meaning lower sleep quality, was related to higher stress and decreased sleep and immobile time. Sleep fragmentation was also positively correlated with greater light exposure on the same day, including red, green and blue light with blue light having the greatest positive correlation with fragmented sleep as shown in previous research (See Fig. 4) [18, 19].

## 4   Discussion

In this study, we investigated the individual and environmental factors that may relate to workplace performance, health and wellbeing as measured by ground truth subjective surveys. Correlations presented in this paper are preliminary due to small sample sizes, however this data is useful in identifying possible connections between the signals and self-reports that can be investigated in computational models. We found two important results in this pilot study: First, that even with seemingly motivated participants, the ability to rely on compliance without feedback from the experiment team can result in inconsistent data across sensors and self-report surveys. Second, that the type of data we are recording can be significantly altered by a few individuals that have drastically different physiology or unexpected activities during a day that can throw off any correlations that may otherwise exist. For larger studies it will be important to be proactive in addressing these concerns first by assisting the participants to improve data collection by giving them feedback on a weekly basis on their data
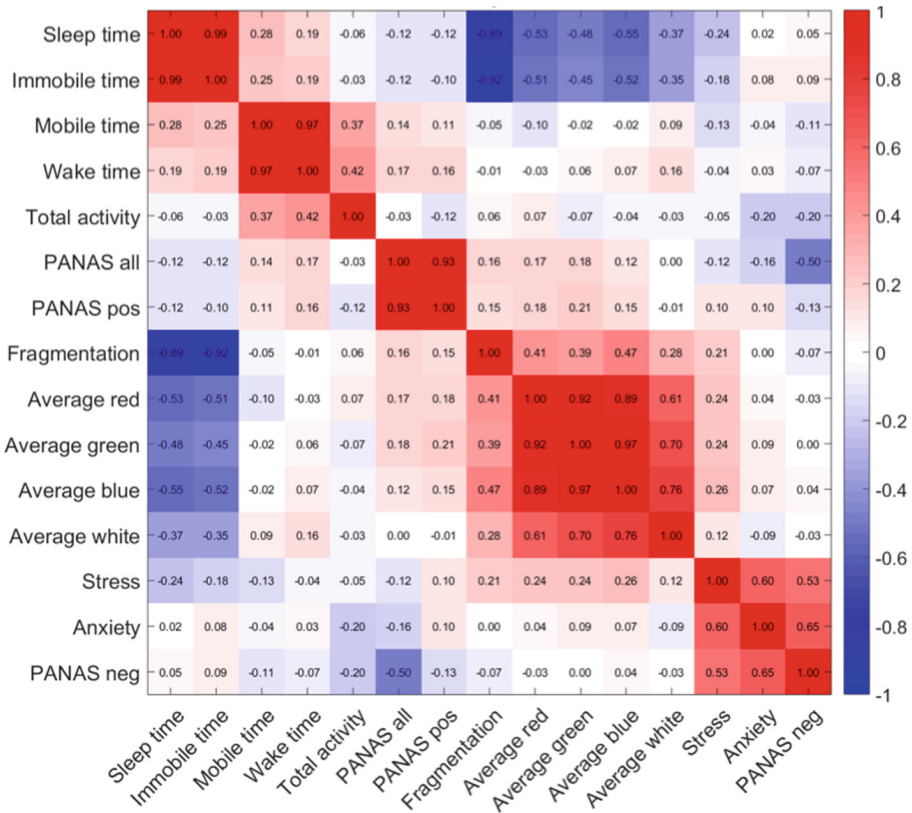
**Fig. 4.** Correlations between DGTB and Actiwatch daily averages

quality and secondly by using a baseline assessment of each individual throughout the experiment to compare changes over time and personalize the models.

Our future plans are to use a dual-track modeling approach to compare workplace performance variables with the data received from this suite of sensors. The theory-driven modeling component of this approach will allow us to base our predictions on well-established and interpretable relationships between the inputs (signal features) and outputs (job performance and wellbeing), while our data-driven modeling component will investigate novel relationships between signals and job performance metrics not yet represented in current literature. Our theory-driven model will have a fast run time with low computational burden. It will also incorporate new connections between signals and individual variables discovered during the data-driven modeling process.

We have shown in prior work that data-driven models, namely neural networks, can be used to discover new connections by removing constraints between signals and individual variables. As the data-driven approach will include many non-linear combinations of signals that are not easily explainable, we will limit our scope of newly discovered links to those that can be identified by our sensitivity analysis techniques [20].

While the methods we use to derive the data-driven model initially require more computational power, adding newly identified connections to the theory-driven model adds only marginal time and burden to the automated process. We can repeat this technique to add novel sensors and automatically derive new connections between signals and variables to further improve the theory-driven model.

The results of this study and our similar ongoing, larger scale study can lead to a better understanding of employee performance, health and wellbeing in the workforce. Future research should investigate the best methods for using this information to positively impact the workplace. A supervisor, worker, or project team could use the results to assess important workplace conditions with reliable and objective metrics. Future applications of these results could allow individuals and groups to re-engineer their workplace processes so as to enhance performance and productivity while lowering workplace stress.

# References

1. Jacobs, J.A., Green, K.: Who are the overworked Americans? Rev. Soc. Econ. **56**(4), 442–459 (1998)
2. Nbm Alberdi, A., Aztiria, A., Basarab, A.: Towards an automatic early stress recognition system for office environments based on multimodal measurements: a review. J. Biomed. Inform. **59**, 49–75 (2016)
3. Pöyhönen, M., Syväoja, S., Hartikainen, J., Ruokonen, E., Takala, J.: The effect of carbon dioxide, respiratory rate and tidal volume on human heart rate variability. Acta Anaesthesiol. Scand. **48**(1), 93–101 (2004)
4. Allen, J.G., MacNaughton, P., Satish, U., Santanam, S., Vallarino, J., Spengler, J.D.: Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office environments. Environ. Health Perspect. **124**(6), 805 (2016). (Online)
5. Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., Gillinov, M.: Accuracy of wrist-worn heart rate monitors. JAMA Cardiol. **2**(1), 104–106 (2017)
6. Sliwka, U., Krasney, J.A., Simon, S.G., et al.: Effects of sustained low-level elevations of carbon dioxide on cerebral blood flow and autoregulation of the intracerebral arteries in humans. Aviat. Space Environ. Med. **69**(3), 299–306 (1998)
7. Pilcher, J.J., Ginter, D.R., Sadowsky, B.: Sleep quality versus sleep quantity: relationships between sleep and measures of health, well-being and sleepiness in college students. J. Psychosom. Res. **42**(6), 583–596 (1997)
8. Figueiro, M.G., Plitnick, B., Rea, M.S.: Light modulates leptin and ghrelin in sleep-restricted adults. Int. J. Endocrinol. **2012** (2012)

9. Vela-Bueno, A., Moreno-Jiménez, B., Rodríguez-Muñoz, A., Olavarrieta-Bernardino, S., Fernández-Mendoza, J., De la Cruz-Troca, J.J., Bixler, E.O., Vgontzas, A.N.: Insomnia and sleep quality among primary care physicians with low and high burnout levels. J. Psychosom. Res. **64**(4), 435–442 (2008)

10. Jackson, M.L., Gunzelmann, G., Whitney, P., Hinson, J.M., Belenky, G., Rabat, A., Van Dongen, H.P.: Deconstructing and reconstructing cognitive performance in sleep deprivation. Sleep Med. Rev. **17**(3), 215–225 (2013)

11. McEwen, B.S.: Sleep deprivation as a neurobiologic and physiologic stressor: allostasis and allostatic load. Metab.-Clin. Exp. **55**, S20–S23 (2006)

12. Eastman, C.I., Young, M.A., Fogg, L.F., Liu, L., Meaden, P.M.: Bright light treatment of winter depression: a placebo-controlled trial. Arch. Gen. Psychiatry **55**(10), 883–889 (1998)

13. Terman, J.S., Terman, M., Schlager, D., Rafferty, B., Rosofsky, M., Link, M.J., Quitkin, F. M.: Efficacy of brief, intense light exposure for treatment of winter depression. Psychopharmacol. Bull. (1990)

14. Gupta, J.K., Lin, C.H., Chen, Q.: Characterizing exhaled airflow from breathing and talking. Indoor Air **20**(1), 31–39 (2010)

15. Puddey, I.B., Beilin, L.J., Vandongen, R., Rouse, I.L., Rogers, P.: Evidence for a direct effect of alcohol consumption on blood pressure in normotensive men. A randomized controlled trial. Hypertension **7**(5), 707–713 (1985)

16. Wasserman, K., Whipp, B.J., Koyl, S.N., Beaver, W.L.: Anaerobic threshold and respiratory gas exchange during exercise. J. Appl. Physiol. **35**(2), 236–243 (1973)

17. Woods, S.W., Charney, D.S., Goodman, W.K., Heninger, G.R.: Carbon dioxide—induced anxiety: behavioral, physiologic, and biochemical effects of carbon dioxide in patients with panic disorders and healthy subjects. Arch. Gen. Psychiatry **45**(1), 43–52 (1988)

18. Chellappa, S.L., Steiner, R., Oelhafen, P., Lang, D., Götz, T., Krebs, J., Cajochen, C.: Acute exposure to evening blue-enriched light impacts on human sleep. J. Sleep Res. **22**(5), 573–580 (2013)

19. Mottram, V., Middleton, B., Williams, P., Arendt, J.: The impact of bright artificial white and 'blue-enriched' light on sleep and circadian phase during the polar winter. J. Sleep Res. **20**(1pt2), 154–161 (2011)

20. Ziegler, M.D., et al.: Sensing and assessing cognitive workload across multiple tasks. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) AC 2016. LNCS (LNAI), vol. 9743, pp. 440–450. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39955-3_41