International Journal of Bioelectromagnetism *www.ijbem.org*
Vol. 22, No. 2, pp. 8 - 21, December 2020.

Wearable Devices for Neurophysiological Evaluation during Real Working-like Tasks: a Reliability Study

Vincenzo Ronca* [a,b] , Alessia Vozzi[a,b], Nicolina Sciaraffa[b,c,d], Antonello Di Florio[b], Luca Tamborra[a,b,e], Ilaria Simonetti[a,b,e], Andrea Giorgi [b], Pietro Aricò[b,c,d], Gianluca Di Flumeri[b,c,d], Fabio Babiloni[b,c,f], and Gianluca Borghini[b,c,d]

[a]*Department of Anatomical, Histological, Forensic and Orthopaedic Sciences, Sapienza University, Rome, Italy*
[b]*BrainSigns srl, Rome, Italy*
[c]*Department of Molecular Medicine, Sapienza University of Rome, Italy*
[d]*IRCCS Fondazione Santa Lucia, Rome, Italy*
[e]*Ernst & Young, Department People Advisory Services, Rome, Italy*
[f]*Department of Computer Science, Hangzhou Dianzi University, Hangzhou, China*

* Corresponding Author: Vincenzo Ronca, Department of Anatomical, Histological, Forensic and Orthopaedic Sciences, Sapienza University, Rome, Italy,
via Regina Elena, 294, Rome, Italy
E-mail: *vincenzo.ronca@uniroma1.it* Phone: +39 3204657112

Abstract. Nowadays, human errors in working environments are among the significant causes of non-fatal and fatal work-related accidents. Therefore, the worker's monitoring and, in particular, the Human Factors (HFs) evaluation plays a crucial role in preventing work-related accidents and in increasing safety in working environments. The present study aimed at assessing the reliability of a specific set of wearable devices, the Empatica E4 and the Muse 2, in the Eye Blinks Rate (EBR), Skin Conductance Level (SCL), and Heart Rate (HR) estimation while performing real working-like tasks. In scientific literature, it is largely demonstrated the effectiveness of monitoring such HFs through specific neurophysiological signals' evaluation. Still, it is also known as the traditional laboratory sensors that imply a certain grade of invasiveness, which could negatively interfere with the worker's performance. The results demonstrated that the wearable devices are reliable as the laboratory technologies for the EBR, SCL, and HR estimation, especially when the time resolution is between 1 and 3 minutes, confirming the possibility of HFs evaluation interfering at minimum the worker's activities.

Keywords: Neurophysiological, Human Factor, Workers, Eye Blinks Rate, Skin Conductance Level, Heart Rate, Wearable

Introduction

According to several European reports, millions of non-fatal and thousands of fatal work-related accidents occurred in the last years (*Quality of Life, quality of public services, and quality of society*, 2016). Such work-related accidents are often caused by Human Factors (HFs) (Hansen, 2006). In this regard, it has been largely demonstrated that high mental workload, tiredness, and stress, cause human errors in different working environments (Bevilacqua & Ciarapica, 2018; Jahangiri, Hoboubi, Rostamabadi, Keshavarzi, & Hosseini, 2016; Melchior & Zanini, 2019; Roets & Christiaens, 2019). These findings clearly indicate the relevance of monitoring in real-time workers' psychophysical state in working operational environments (Mehta & Parasuraman, 2013). In this context, scientific literature largely highlighted the limitations of using subjective methodologies to evaluate such HFs (Aricò et al., 2017; Babiloni, 2019a; Wall et al., 2004). As a potential countermeasure, in the last decades, neuroscientific disciplines have been dedicating a consistent effort in investigating human physiological correlates of user's mental states to develop monitoring tools able to detect incoming cognitive impairments (e.g., mental overload) based on specific biomarkers (e.g., skin sweating fluctuations, heart rate variability, brain electrical activity variations in specific rhythms over particular cortical sites)

(Sciaraffa et al., 2017; Aricó et al., 2017; Borghini, Aricò, Di Flumeri, et al., 2017; Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014b; Fairclough, 2009). Nevertheless, the need to reduce the invasiveness related to monitoring methodologies during the working tasks to do not negatively interfere with the workers' activities and comfort (Charles & Nixon, 2019). This last consideration is very consistent with the concept of monitoring by the mean of wearable devices. Wearable development progressively increased in the industrial and scientific world (Ragot, Martin, Em, Pallamin, & Diverrez, 2018). Such systems generally collect data and send them wirelessly to a central computation unit. Therefore, wearable devices are fully compatible with unobtrusive workers' monitoring. Such sensors can continuously track the worker's mental state without negatively interfering with its daily working activities. On the other hand, several doubts could emerge about wearable devices' reliability and capability concerning Laboratory equipment. Usually, wearable devices employ dry electrodes, and they integrate fewer sensors than the most reliable laboratory sensors. Laboratory devices are very effective (Ragot et al., 2018), but they cannot be used in realistic settings in most cases. Different previous works already explored the comparison between wearable and laboratory devices. For example, Ragot and colleagues (Ragot et al., 2018) demonstrated the reliability of emotion recognition outside the laboratory using wearable devices. Pang and colleagues (Pang, Okubo, Sturnieks, Lord, & Brodie, 2019) showed the effectiveness of wearable devices support in older people to prevent physical injuries, while the laboratory sensors were not compatible with daily usage. Fuller and colleagues (Fuller et al., 2020) proved the robustness of commercial wearable devices for tracking daily activities, such as the daily steps and the user's heart rate along the day.

The present study aimed at assessing the reliability of the wearable devices with respect to the laboratory ones for their further use in working environments. The abovementioned works did not consider this field of application. In particular, in this study, specific parameters estimated from the Galvanic Skin Response (GSR), Photoplethysmographic (PPG), and Electrooculographic (EOG) signals collected by the wearable sensors were compared with the corresponding ones estimated from the GSR, Electrocardiographic (ECG), and EOG signals gathered by the laboratory equipment. Such neurophysiological signals were selected from the future perspective of characterizing specific user's cognitive and emotional states while dealing with different operative tasks (Babiloni, 2019b). In scientific literature, it has been largely demonstrated that the abovementioned neurophysiological parameters play a crucial role in the mental workload, stress and tiredness evaluation (Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014a; Borghini, Ronca, et al., 2020).

Material and Methods

1.1 Participants

Seventeen (17) participants from the Sapienza University of Rome, ten males and seven females (31.1 ± 3.7 years old), were recruited and involved voluntarily in this study. Informed consent was obtained from each subject after explaining the study, approved by the local institutional ethics committee. The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000. It was approved by the Sapienza University of Rome Ethical Committee in Charge for the Department of Molecular Medicine. The participants were instructed about the tasks' and questionnaires' execution. Then, they were asked to wear both wearable and laboratory devices. In particular, they wore the headband Muse 2 (InteraXon Inc., Canada) on the forehead, the smartband Empatica E4 (Empatica, Italy) on the wrist of their non-dominant hand, the Shimmer3 GSR+ (Shimmer Sensing, Ireland) was set on their index and middle finger of the same hand, one gel-based Ag/AgCl electrode was set on the participant's Fpz scalp location, and an additional gel-based electrode was placed on the participant's chest, both of them were connected to the BEMicro system (EBNeuro, Italy) together with other EEG electrodes not employed for the purposes of the present study. These gel-based electrodes were considered for the Heart Rate (HR) and the Eye Blink Rate (EBR) evaluation, respectively. A picture representing the final setting is shown in Figure 1.

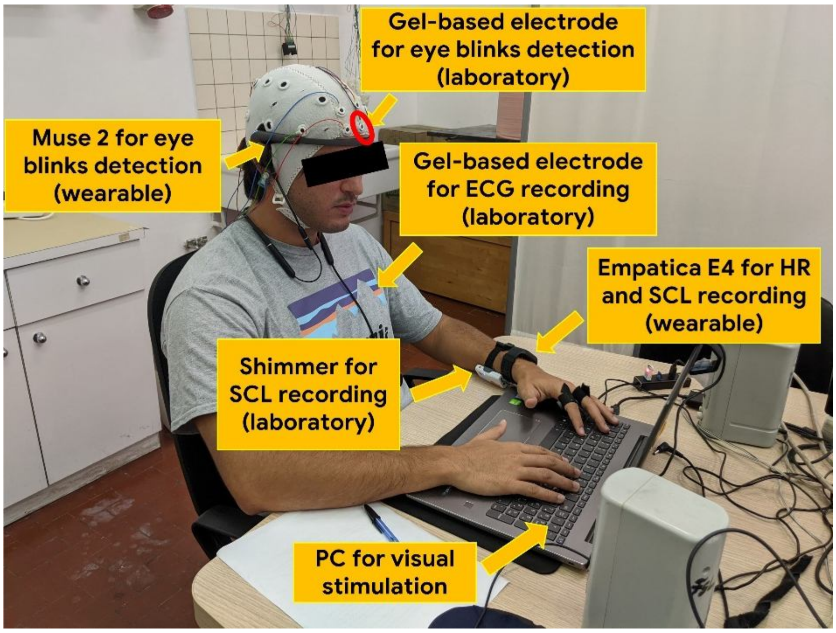

Figure 1*. The final setting of the laboratory and wearable sensors.*

1.2 Experimental protocol

The experimental protocol was designed to represent three experimental tasks (Figure 2) designed to simulate three specific real working environments: the office, the teleworking, and the manufacturing use case. A detailed description of such tasks is provided below:

- The *Office-like* workstation was reproduced using the n-Back task. The n-Back (NB) task is a well-known computer-based psychological test used to manipulate workload, or more specifically, working memory load ("Age differences in short-term retention of rapidly changing information. - PsycNET," n.d.; Kane, Conway, Miura, & Colflesh, 2007). Memory load is considered a major component and reasonable approximation of workload (Berka et al., n.d.). Within this task, a sequence of stimuli is presented to the test person. The task is to indicate when the current stimulus matches the stimulus that occurred in the series n steps before. The factor n can be adjusted to make the task more difficult or easier. A baseline and three conditions of such task were tested in the proposed study, two of them with different difficulty levels and one stressful situation. In all conditions, 21 uppercase letters were used, which were displayed for 500 ms and an inter-stimulus interval randomized between 500 to 3000ms; 33% of the displayed letters were the target.
- To simulate the *Teleworking* Use Case, two interactive web calls (WE) were performed. Microsoft Teams software (Microsoft, USA) was used for the interactive web calls. Three conditions of such task were tested: i) Baseline condition, in which the participants looked at the Microsoft Teams user interface without reacting; ii) Positive condition, in which the test persons were asked to report the happiest memory of their life; iii) Negative condition, in which the test persons were asked to report the saddest memory of their life.
- The *Manufacturing* use case was simulated by a fine motor skills task, the children's game "Doctor Game (DG)." The aim of the game was to remove small objects from the board without touching the edges. Here too, a baseline, two difficulty levels, and one stressful condition were tested. In all conditions, the time required by the test subject to complete the task was measured, and the errors made were noted. The easy and the hard condition were performed twice each.

Furthermore, at the beginning of the experiments, a baseline measurement was taken. The participant was asked to look calmly straight ahead, one minute with open eyes and one minute with closed eyes. The three tasks and the task conditions were randomized across the participants to avoid both habituation and expectation effects. After every stressful condition and also after the web call in which the negative experience was asked for, the test person had 4 minutes to regenerate, and relaxing music was played to support him/her.

Such experimental tasks were designed to modulate the subjects' mental workload, stress, and emotional state. However, this aspect will not be investigated in the proposed study.

The present study's primary objective consisted of evaluating the capability and reliability of the wearable technology in estimating the GSR, HR, and EBR features. In particular, such parameters were evaluated with a time resolution of 60 seconds and by averaging the features within each experimental condition, which had a duration from one to three minutes, to assess their robustness associated with these two temporal resolutions.

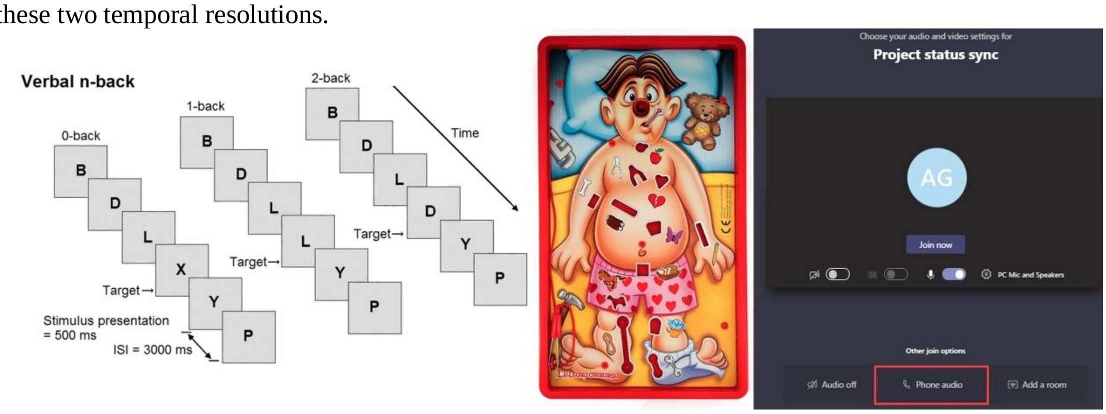

Figure 2*. Examples of the tasks included in the experimental protocol: the n-Back, Doctor Game, and Web call task.*

Data Collection and Data Processing

1.3 Eyeblink signal recording and analysis

The EOG signals were acquired simultaneously by the Fpz channel of the BeMicro and TP9 channel of the Muse 2 and with a sampling frequency of 256 (Hz) and 64 (Hz), respectively. Regarding Muse 2, the TP9 channel was selected because it features the reference electrode in Fpz scalp location. Therefore, to identify the eye blinks, which are generally identifiable in Fpz scalp location, we analyzed the signal recorded by the TP9 channel. Such a channel includes the subtraction of the signal captured by Fpz channel.

The EOG analysis aimed to detect eye blinks to estimate the Eye Blinks Rate (EBR). Both the dataset were analyzed by using the same algorithm. Firstly, the EOG signal has been band-pass filtered using a 201^{st} order Butterworth filter within the frequency range of 2- 10 (Hz). The eye blinks detection method was performed in two main steps:

1) Threshold calculation

2) Pattern Matching

In (1), the Eyes Open condition was used to identify a threshold that, when exceeded, identified a potential blink. The threshold was calculated as follows:

$$Threshold = mean(EEG\ Eyes\ Open) + 3 * robustStdDev$$

where $std\_threshold = 3$, while $robustStdDev$ is the mean absolute deviation of the corresponding EOG channel.

In (2), every time the EOG signal exceeded the computed threshold, the Pearson correlation between a common blink template and the EOG signal was computed within each experimental condition. If this value was higher than 0.9, a potential blink would be classified as a “real blink.”

The EBR feature estimated for each participant in each condition was calculated as the mean of the total number of blinks in every condition per minute.

1.4 GSR signal recording and analysis

The GSR was recorded by both a Laboratory and Wearable device placed on the participant’s non-dominant hand. The Shimmer3 GSR+ unit, a Laboratory device, acquired the GSR with a sampling frequency of 64 (Hz) using two electrodes on the index and middle fingers. The Empatica E4 acquired the GSR with a sampling frequency of 4 (Hz) using the two electrodes placed on the bottom part of the watch strap.

The GSR was first low-pass filtered with a cut-off frequency of 1 (Hz) and then processed using the *Ledalab* suite (Bach, 2014), a specific opensource toolbox implemented within MATLAB GSR processing. *The continuous decomposition analysis* (Benedek & Kaernbach, 2010) as applied to estimate the *tonic* (SCL) and the *phasic* (SCR) components (Braithwaite, Derrick, Watson, Jones, & Rowe, n.d.; Posada-Quintero, Florian, Orjuela-Cañón, & Chon, 2018). The SCL is the slow-changing part of the GSR signal, mostly related to the participant's global arousal. The SCR is the fast-changing part of the GSR signal, which occurs in relation to single stimuli reactions. The GSR components (and the other neurophysiological parameters) were estimated with a 60 seconds time resolution and as the mean within each experimental condition. Finally, only the SCL component was considered for the participant's stress evaluation, as demonstrated by Borghini et al.(Borghini, Di Flumeri, et al., 2020).

1.5 ECG signal recording and analysis

The Electrocardiogram (ECG) was recorded using a Laboratory and Wearable technology for the abovementioned signals. In particular, the participants' ECG was recorded using an electrode fixed on their chest (Laboratory device) and referred to the potential recorded at both the earlobes, with a sampling frequency of 256 Hz. Simultaneously, the Empatica E4 acquired the Photoplethysmography (PPG) with a sampling frequency of 64 (Hz) the sensors placed in the bottom part of the watch. First, the ECG and PPG signals were filtered using a 5th-order Butterworth band-pass filter (1 – 15 Hz, and 1 – 4 Hz, respectively) to reject the continuous component and the high-frequency interferences, such as that related to the mains power source. At the same time, the purpose of this filtering was to emphasize the QRS process of the ECG signal (Goovaerts, Ros, van den Akker, & Schneider, 1976; Thakor, Webster, & Tompkins, 1980). The following step consisted of computing the ECG (PPG) signal to the power of 3 to emphasize the heartbeat peaks, as they generally have the higher amplitude, and at the same time, reduce spurious artifact peaks. Finally, we measured the distance between consecutive peaks (i.e., each R peak corresponds to a heartbeat) to estimate the *Heart Rate* (HR) values every 60 seconds.

All the parameters mentioned above, i.e., the EBR, SCL, and HR, were finally normalized to obtain comparable distributions related to each sensor technology employed in the study. The normalization consisted in the subtraction of the baselines from the respective values estimated during each experimental condition.

Results

1.6 Eye blinks analysis results

Regarding the eye blinks parameters estimation, the paired Wilcoxon signed-rank test performed on the normalized EBR within all the experimental tasks did not show any significant difference between the wearable and the laboratory technology ($p = 0.4$). Such aspect was also confirmed by the graph reported in Figure 3, where the differences in terms of EBR estimation(y-axis) between the wearable and laboratory technology are reported for each participant (x-axis).

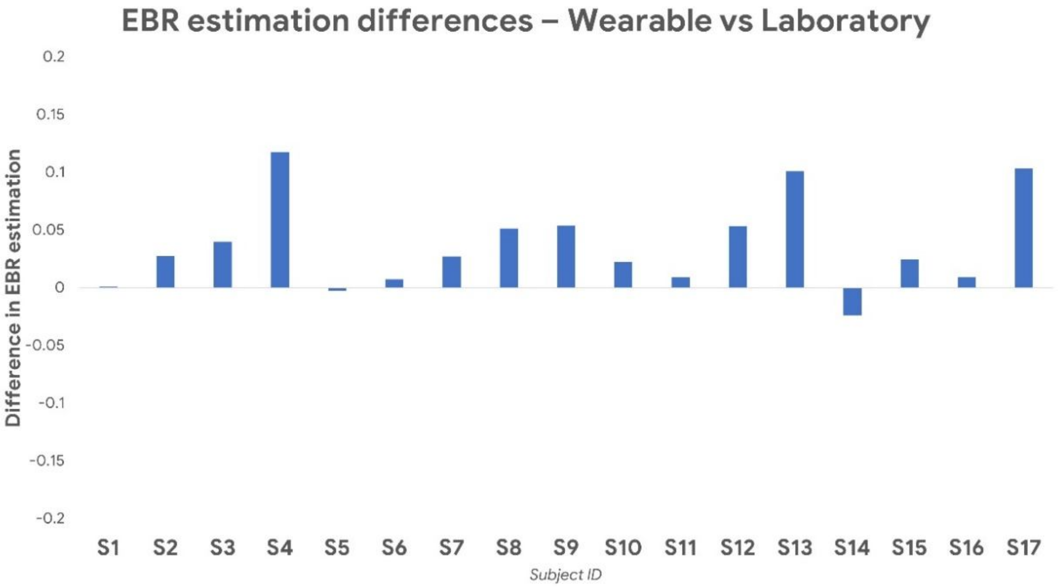

Figure 3*. EBR estimation differences between the values estimated respectively by the wearable and laboratory devices for each participant (x-axis). It can be observed that only for three subjects, the difference in EBR estimation was higher than 0.1.*

To further investigate the wearable devices' capability with respect to the laboratory ones, two different repeated measure correlation analyses were performed. The first analysis aimed to compare the estimation of the parameters every 60 seconds, while in the second one, we averaged the EBR estimations within each experimental condition to obtain a lower time resolution ranging between 1 and 3 minutes. As reported in Figure 4, the repeated measure correlation analysis (Bakdash & Marusich, 2017) performed between the EBR estimated by the laboratory and wearable device every 60 seconds showed a positive (R = 0.8) and significant ($p < 10^{-64}$) correlation, a demonstration of how the two devices provided similar EBR estimations.

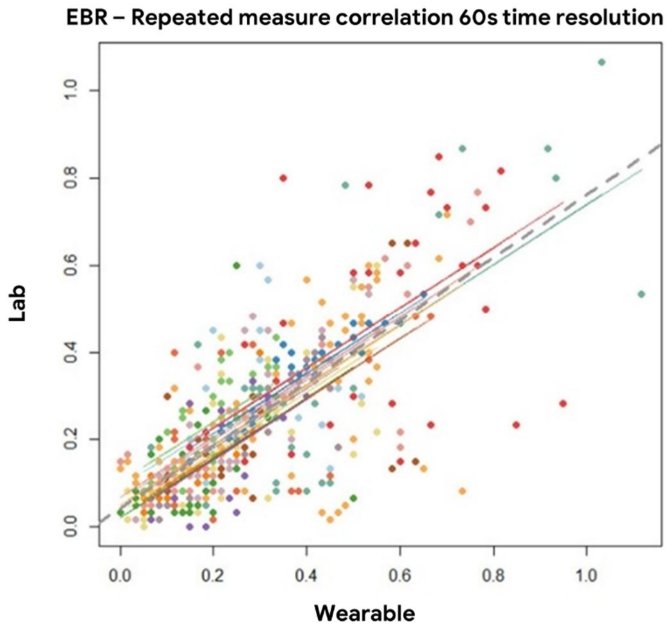

Figure 4*. The repeated measure correlation analysis on the EBR estimated by the laboratory and wearable devices every 60 seconds.*

Similarly, the repeated measure correlation analysis was performed on the EBR estimation within each experimental condition, obtaining a lower time resolution ranging between 1 and 3 minutes. The results reported in Figure 5 confirmed how the estimations of the EBR provided by the wearable and laboratory devices were highly correlated ($R = 0.9$, $p = < 10^{-123}$) .

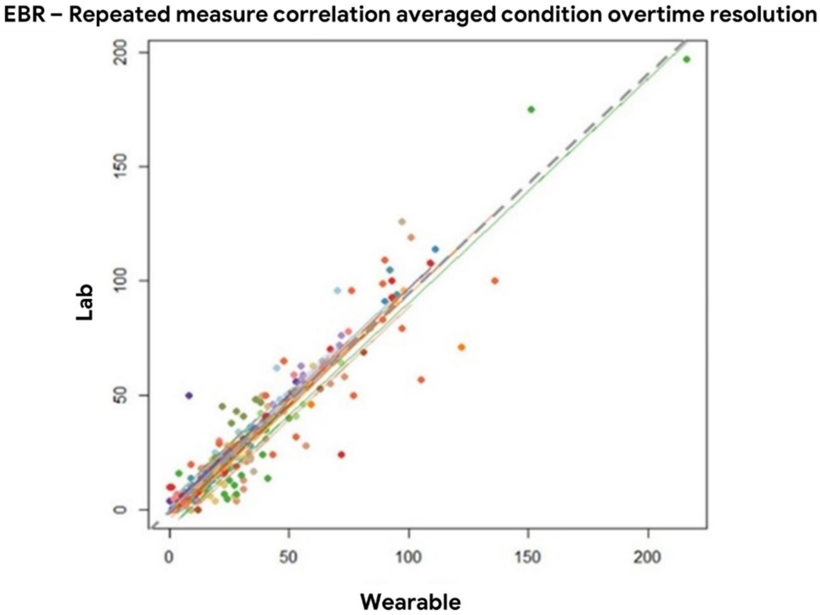

Figure 5*. The repeated measure correlation analysis on the EBR estimated by the laboratory and wearable devices within each experimental condition.*

1.7 SCL analysis results

Regarding the SCL parameter estimation, the paired Wilcoxon signed-rank test performed on the normalized SCL within the three experimental tasks revealed a non-significant difference between the wearable and the laboratory technology ($p = 0.2$). Such evidence can also be observed in the qualitative analysis performed on SCL estimations provided by the wearable and laboratory sensors. The results showed that the difference in SCL estimation between the values evaluated by the wearable and laboratory sensors was within a range of ± 3 µS for most participants (Figure 6).

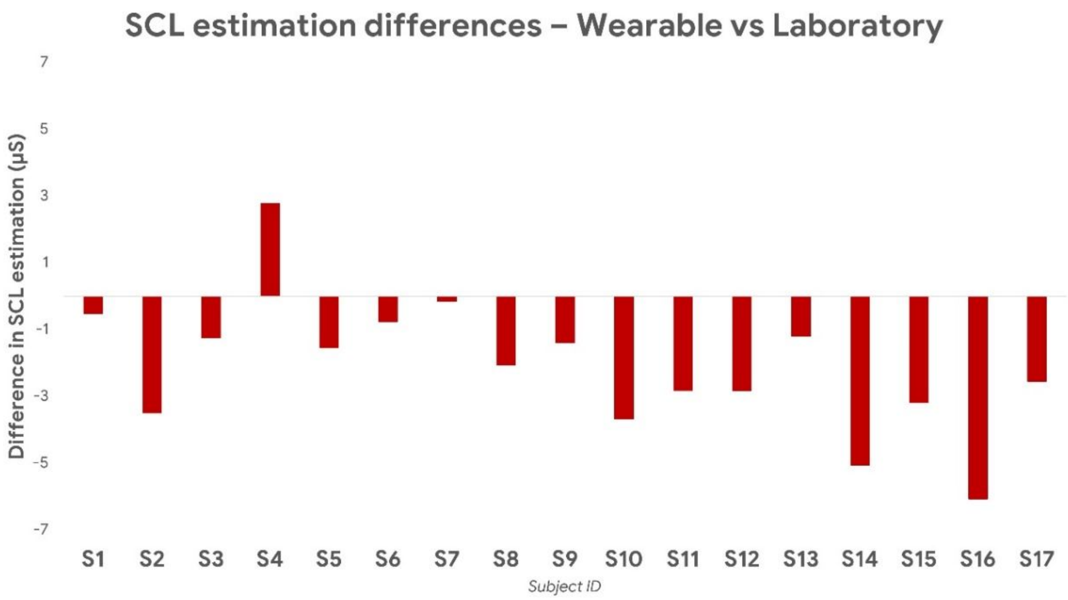

Figure 6*. Differences in SCL estimation between the values estimated respectively by the wearable and laboratory devices for each participant (x-axis).*

The repeated measure correlation analysis performed between the SCL estimated by the laboratory and wearable devices every 60 seconds did not show an overall significant correlation (R = 0.3, p = < 10^{-10}) (Figure 7). For some participants, the correlation was low, whereas, for others, it was very high.

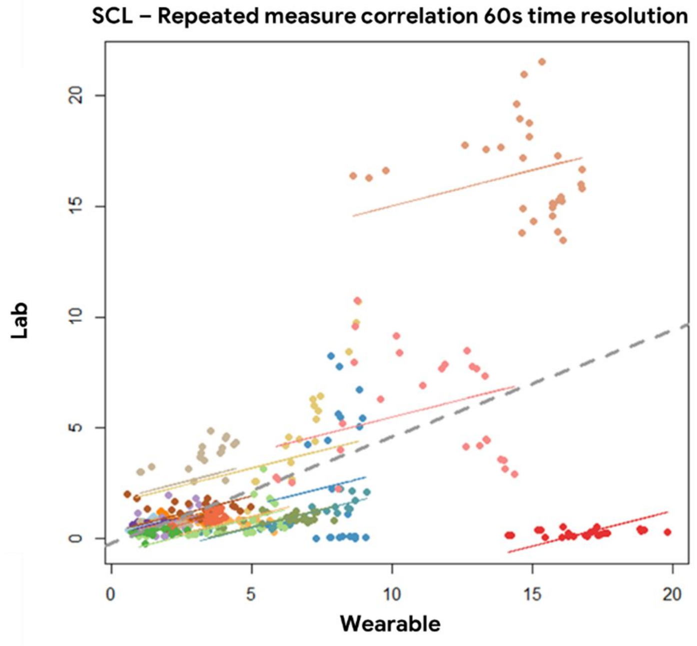

***Figure 7**. The repeated measure correlation analysis on the SCL is estimated by the laboratory and wearable devices every 60 seconds.*

A different discussion can be done for the results from the repeated measure correlation analysis performed with a lower temporal resolution (Figure 8), which was computed by averaging the SCL values within each experimental condition. Both the analyses showed positive and significant (R = 0.5, p = < 10^{-9}) correlations between the SCL estimations of the two technologies. These results demonstrated that with a lower time resolution, the SCLs estimated by the wearable device showed more similar variations to the ones estimated by the laboratory device.

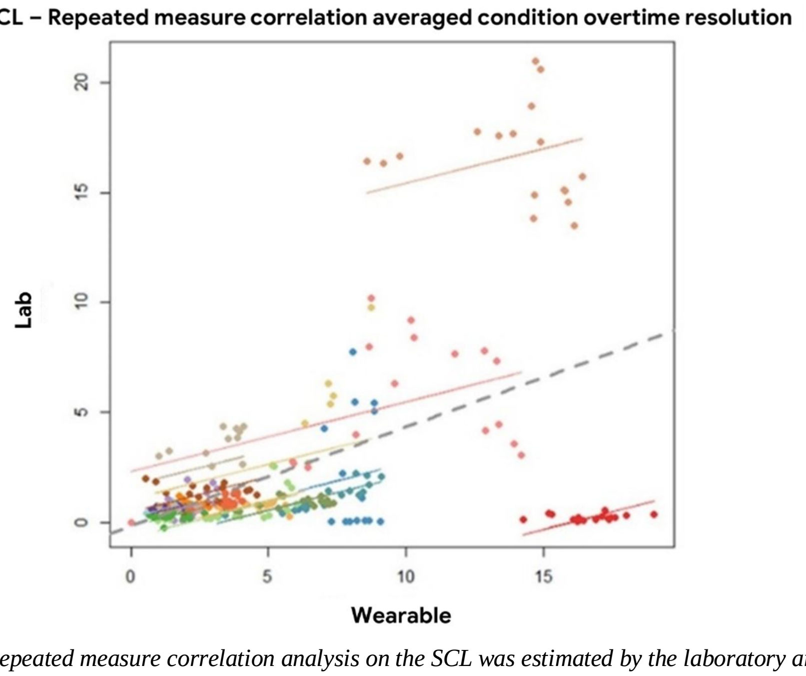

***Figure 8**. The repeated measure correlation analysis on the SCL was estimated by the laboratory and wearable devices within each experimental condition.*

1.8 HR analysis results

The Wilcoxon signed-rank test on the normalized HR within the three experimental tasks did not report any significant differences ($p = 0.4$) between the two technologies to demonstrate that the HR's absolute values were also similar. In other words, there was no difference in estimating HR via wearable or laboratory technology. Such evidence can also be observed in Figure 9, where all the differences in HR estimation provided by the wearable and laboratory devices were within ± 6 BPM.

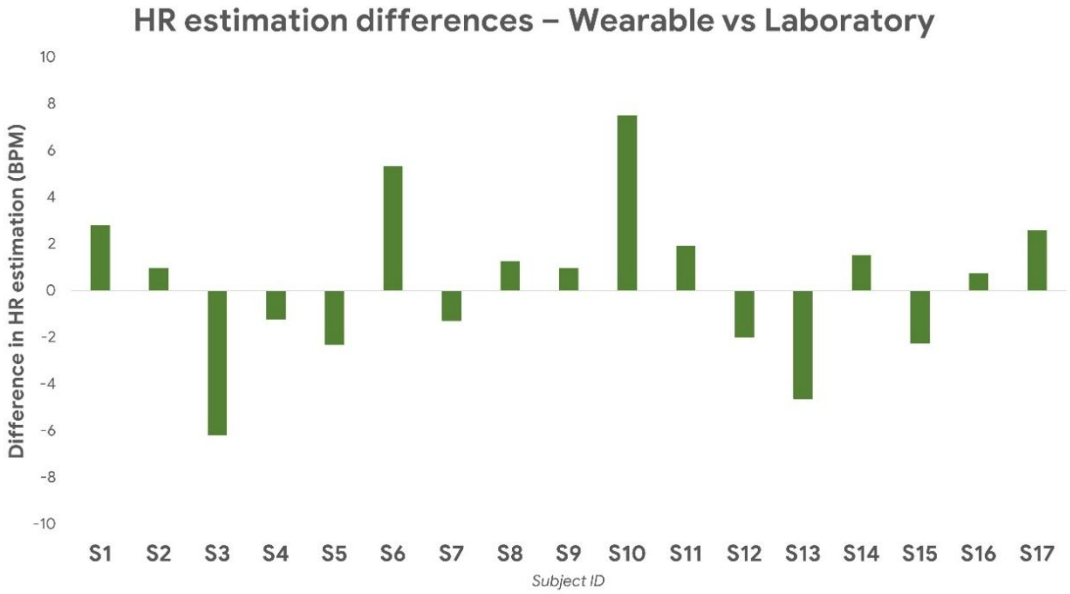

Figure 9*. Differences in HR estimation between the values estimated respectively by the wearable and laboratory devices for each participant (x-axis)..*

As for the EBR and SCL parameters, the repeated measure correlation analysis was performed on the HR parameter. The results (Figure 10 and Figure 11) demonstrated that the HR values estimated by the wearable, either every 60 seconds or within each experimental condition, are significantly correlated ($R = 0.5$, $p = < 10^{-15}$) to the one estimated by the laboratory technology. In particular, as found for the SCL component, when the HR was estimated with a temporal resolution ranging between 1 and 3 minutes, the correlation between the laboratory and wearable technology (Figure 11) was overall high and significant ($R = 0.7$, $p < 10^{-20}$).

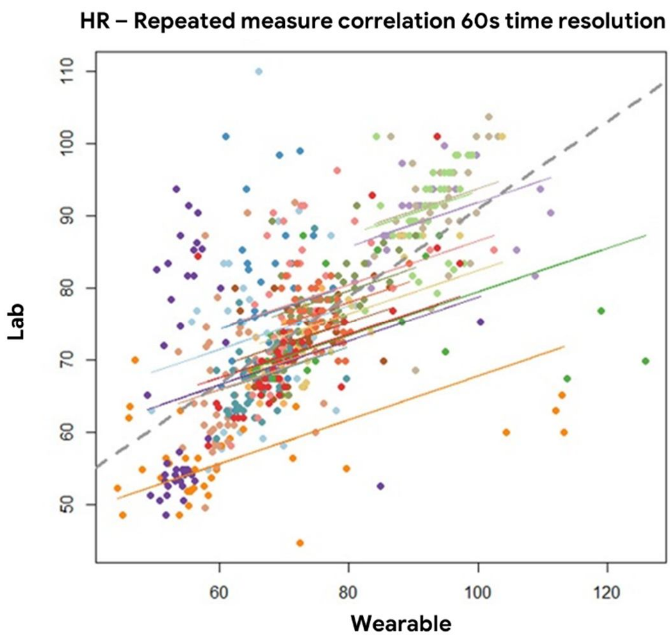

Figure 10*. The repeated measure correlation analysis on HR is estimated by the laboratory and wearable devices every 60 seconds.*

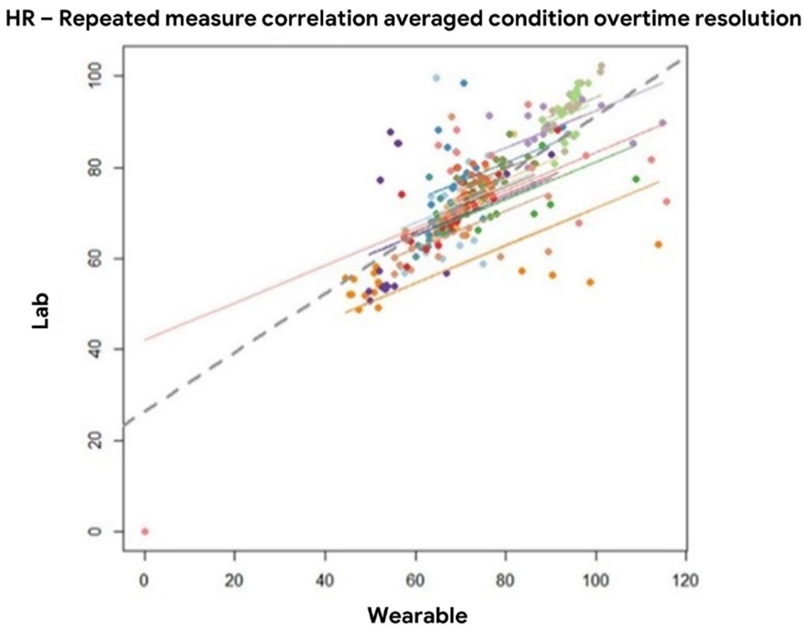

Figure 11*. The repeated measure correlation analysis on HR was estimated by the laboratory and wearable devices within each experimental condition.*

Discussion

The present work aimed at investigating the reliability of wearable technologies in estimating EBR, SCL and HR while performing working-like tasks, with respect to laboratory equipment. In this regard, two main aspects emerged from the results supporting the reliability of the wearable devices. On the one hand, the statistics did not show any significant difference between the values estimated by the wearable devices and the ones evaluated through the laboratory sensors for all the three parameters considered. On the other hand, the correlations, particularly those performed on the parameters averaged within each experimental condition, by considering a lower temporal resolution, between the wearable and the laboratory technology were high and significant. Such two aspects confirm that, compared to the laboratory technologies, the wearables could capture the parameters considered' dynamics with the same capability offered by the laboratory devices.

More specifically, the SCL and the HR estimated by the Empatica E4 did not differ from the ones evaluated through the Shimmer3 GSR+, while, the Muse 2 was able to efficiently estimate the EBR as done through the EOG channel. In fact, the EBR estimations of the laboratory and wearable technology were the same either if done every 60 seconds or by averaging the EBR values within each experimental condition, i.e., each time interval ranging between 1 and 3 minutes. Such results are coherent with the evidence proposed by the previous related works. As showed in the Ragot and colleagues' study (Ragot et al., 2018), the GSR and HR measurements provided by the Empatica E4 wristband resulted robust as the one provided by the laboratory sensors; while the Muse 2 demonstrated the same reliability as the gel-based electrode placed in Fpz scalp location in terms of EBR estimation.

Although promising and interesting results, there are some limitations to be discussed. The wearable devices imply specific technical constraints, such as the limited battery life since such devices operate wirelessly. Moreover, the correlations between the parameters evaluated through the wearable devices and the ones measured through the laboratory devices increased significantly (above 0.7) only when the indexes were evaluated every 2 – 3 minutes. This aspect could prevent the real time monitoring at high temporal resolution through wearable technologies. The proposed study was a preliminary exploration of wearable reliability. In this regard, a next experimental protocol has been planned in which the discrimination of mental workload, stress and emotional state through the wearable devices will be investigated. Moreover, such a next experimental protocol will include longer experimental condition, in order to investigate the wearables' reliability with different combinations of time resolution. In addition, such a next experimental protocol will allow us to better investigate the influence of subject's movements, in terms of typology and frequency of movements, on the wearables reliability, since we observed that this dry-electrode based technology could more easily lose contact with the skin compared with the gel-based electrode technology.

Conclusions

The analysis and comparison between the laboratory (the gel-based electrodes connected to the BEMicro system and the GSR+ Shimmer3) and wearable (the Muse 2 and the Empatica E4) technology demonstrated that the latter could be used for reliably estimating the participant's EBR, HR, and SCL. In general, we can conclude that the wearable devices are reliable as the laboratory ones, especially when the time resolution is about 2 – 3 minutes.

Besides the reliability, the wearable technologies offer the advantage that they do not require technical personnel for the setup. Such a feature could play a crucial role in experimental protocols including a large sample size simultaneously. Finally, it has to be noted that the worker's monitor through wearable sensors is very promising in different applications, implying the minimum interference with the user's activities, therefore paving the way for future safety-oriented applications (Aricò et al., 2015; Arico et al., 2018; Borghini, Arico, et al., 2020; Borghini, Ronca, et al., 2020; Di Flumeri et al., 2019), other mental states assessment like stress (Borghini, Di Flumeri, et al., 2020), vigilance (Sebastiani et al., 2020), and cognitive control behavior (Wang, Moreau, & Kao, 2019; Borghini, Aricò, DI Flumeri, et al., 2017; Melchior & Zanini, 2019).

Acknowledgments

This work is co-financed by the European Commission by Horizon2020 projects "WORKINGAGE: Smart Working environments for all Ages" (GA n. 826232); "SIMUSAFE: Simulator Of Behavioral Aspects For Safer Transport" (GA n. 723386); "SAFEMODE: Strengthening synergies between Aviation and maritime in the area of human Factors towards achieving more Efficient and resilient MODE of transportation" (GA n. 814961), "BRAINSAFEDRIVE: A Technology to detect Mental States during Drive for improving the Safety of the road" (Italy-Sweden collaboration) with a grant of Ministero dell'Istruzione dell'Università e della Ricerca della Repubblica Italiana, "MINDTOOTH: Wearable device to decode human mind by neurometrics for a new concept of smart interaction with the surrounding environment" (GA n. 950998).

References

Age differences in short-term retention of rapidly changing information. - PsycNET. (n.d.). Retrieved December 1, 2020, from https://psycnet.apa.org/record/1959-07784-001

Aricò, P., Borghini, G., Di Flumeri, G., Bonelli, S., Golfetti, A., Graziani, I., … Babiloni, F. (2017). Human Factors and Neurophysiological Metrics in Air Traffic Control: A Critical Review. *IEEE Reviews in Biomedical Engineering*, Vol. 10, pp. 250–263. https://doi.org/10.1109/RBME.2017.2694142

Aricó, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Colosimo, A., & Babiloni, F. (2017). Passive BCI in operational environments: Insights, recent advances, and future trends. *IEEE Transactions on Biomedical Engineering*, *64*(7), 1431–1436. https://doi.org/10.1109/TBME.2017.2694856

Aricò, P., Borghini, G., Flumeri, G. Di, Colosimo, A., Graziani, I., Imbert, J.-P., … Pozzi, S. (2015). Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks. *IEEE*, 9781424492695. https://doi.org/10.1109/EMBC.2015.7320063ï

Arico, P., Reynal, M., Imbert, J. P., Hurter, C., Borghini, G., Di Flumeri, G., Babiloni, F. (2018). Human-machine interaction assessment by neurophysiological measures: A study on professional air traffic controllers. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2018-July*, 4619–4622. https://doi.org/10.1109/EMBC.2018.8513212

Babiloni, F. (2019a). Mental workload monitoring: New perspectives from neuroscience. *Communications in Computer and Information Science*, *1107*, 3–19. https://doi.org/10.1007/978-3-030-32423-0_1

Babiloni, F. (2019b). Mental workload monitoring: New perspectives from neuroscience. *Communications in Computer and Information Science*, *1107*, 3–19. https://doi.org/10.1007/978-3-030-32423-0_1

Bach, D. R. (2014). A head-to-head comparison of SCRalyze and Ledalab, two model-based methods for skin conductance analysis. *Biological Psychology*, *103*(1), 63–68. https://doi.org/10.1016/j.biopsycho.2014.08.006

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*(MAR), 456. https://doi.org/10.3389/fpsyg.2017.00456

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., … Craven, P. L. (n.d.). *EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks*.

Bevilacqua, M., & Ciarapica, F. E. (2018). Human factor risk management in the process industry: A case study. *Reliability Engineering and System Safety*, *169*, 149–159. https://doi.org/10.1016/j.ress.2017.08.013

Borghini, G., Aricò, P., DI Flumeri, G., Cartocci, G., Colosimo, A., Bonelli, S., … Babiloni, F. (2017). EEG-based cognitive control behaviour assessment: An ecological study with professional air traffic controllers. *Scientific Reports*, *7*(1), 1–16. https://doi.org/10.1038/s41598-017-00633-7

Borghini, G., Aricò, P., Di Flumeri, G., Sciaraffa, N., Colosimo, A., Herrero, M. T., … Babiloni, F. (2017). A new perspective for the training assessment: Machine learning-based neurometric for augmented user's evaluation. *Frontiers in Neuroscience*, *11*(JUN), 325. https://doi.org/10.3389/fnins.2017.00325

Borghini, G., Arico, P., Di Flumeri, G., Sciaraffa, N., Ronca, V., Vozzi, A., & Babiloni, F. (2020). Assessment of athletes' attitude: Physiological evaluation via wearable sensors during grappling competitions. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2020-July*, 584–587. https://doi.org/10.1109/EMBC44109.2020.9176401

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014a). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014b, July 1). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, Vol. 44, pp. 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Borghini, G., Di Flumeri, G., Aricò, P., Sciaraffa, N., Bonelli, S., Ragosta, M., … Babiloni, F. (2020). A multimodal and signals fusion approach for assessing the impact of stressful events on Air Traffic Controllers. *Scientific Reports*, *10*(1), 8600. https://doi.org/10.1038/s41598-020-65610-z

Borghini, G., Ronca, V., Vozzi, A., Aricò, P., Di Flumeri, G., & Babiloni, F. (2020). Monitoring performance of professional and occupational operators. In *Handbook of Clinical Neurology* (Vol. 168, pp. 199–205). https://doi.org/10.1016/B978-0-444-63934-9.00015-9

Braithwaite, J. J., Derrick, D., Watson, G., Jones, R., & Rowe, M. (n.d.). *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*.

Charles, R. L., & Nixon, J. (2019, January 1). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, Vol. 74, pp. 221–232. https://doi.org/10.1016/j.apergo.2018.08.028

Di Flumeri, G., Aricò, P., Borghini, G., Sciaraffa, N., Ronca, V., Vozzi, A., … Babiloni, F. (2019). EEG-Based workload index as a taxonomic tool to evaluate the similarity of different robot-assisted surgery systems. *Communications in Computer and Information Science*, *1107*, 105–117. https://doi.org/10.1007/978-3-030-32423-0_7

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers*, *21*(1–2), 133–145. https://doi.org/10.1016/j.intcom.2008.10.011

Fuller, D., Colwell, E., Low, J., Orychock, K., Ann Tobin, M., Simango, B., … Taylor, N. G. A. (2020, September 1). Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rates Systematic review. *JMIR MHealth and UHealth*, Vol. 8, p. e18694. https://doi.org/10.2196/18694

Goovaerts, H. G., Ros, H. H., van den Akker, T. J., & Schneider, H. (1976). Communications: A digital QRS detector based on the principle of contour limiting. *IEEE Transactions on Biomedical Engineering*, *BME-23*(2), 154–160. https://doi.org/10.1109/TBME.1976.324577

Hansen, F. D. (2006). Human error: A concept analysis. *Journal of Air Transportation*. Retrieved from https://ntrs.nasa.gov/search.jsp?R=20070022530

Jahangiri, M., Hoboubi, N., Rostamabadi, A., Keshavarzi, S., & Hosseini, A. A. (2016). Human error analysis in a permit to work system: A case study in a chemical plant. *Safety and Health at Work*, *7*(1), 6–11. https://doi.org/10.1016/j.shaw.2015.06.002

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(3), 615–622. https://doi.org/10.1037/0278-7393.33.3.615

Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in Human Neuroscience*, *7*(DEC), 889. https://doi.org/10.3389/fnhum.2013.00889

Melchior, C., & Zanini, R. R. (2019, April 1). Mortality per work accident: A literature mapping. *Safety Science*, Vol. 114, pp. 72–78. https://doi.org/10.1016/j.ssci.2019.01.001

Pang, I., Okubo, Y., Sturnieks, D., Lord, S. R., & Brodie, M. A. (2019). Detection of near falls using wearable devices. *Journal of Geriatric Physical Therapy*, *42*(1), 48–56. https://doi.org/10.1519/JPT.0000000000000181

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., & Chon, K. H. (2018). Electrodermal activity is sensitive to cognitive stress under water. *Frontiers in Physiology*, *8*(JAN), 1128. https://doi.org/10.3389/fphys.2017.01128

Quality of life, quality of public services, and quality of society. (2016). https://doi.org/10.2806/707101

Ragot, M., Martin, N., Em, S., Pallamin, N., & Diverrez, J. M. (2018). Emotion recognition using physiological signals: Laboratory vs. wearable sensors. *Advances in Intelligent Systems and Computing*, *608*, 15–22. https://doi.org/10.1007/978-3-319-60639-2_2

Roets, B., & Christiaens, J. (2019, March 4). Shift work, fatigue, and human error: An empirical analysis of railway traffic control. *Journal of Transportation Safety and Security*, Vol. 11, pp. 207–224. https://doi.org/10.1080/19439962.2017.1376022

Sciaraffa, N., Borghini, G., Aricò, P., Di Flumeri, G., Colosimo, A., Bezerianos, A., … Babiloni, F. (2017). Brain interaction during cooperation: Evaluating local properties of multiple-brain network. *Brain Sciences*, *7*(7). https://doi.org/10.3390/brainsci7070090

Sebastiani, M., Di Flumeri, G., Aricò, P., Sciaraffa, N., Babiloni, F., & Borghini, G. (2020). Neurophysiological vigilance characterisation and assessment: Laboratory and realistic validations involving professional air traffic controllers. *Brain Sciences*, *10*(1), 48. https://doi.org/10.3390/brainsci10010048

Thakor, N. V., Webster, J. G., & Tompkins, W. J. (1980, January 1). *Optimal QRS Filter*. Retrieved from https://jhu.pure.elsevier.com/en/publications/optimal-qrs-filter-3

Wall, T. D., Michie, J., Patterson, M., Wood, S. J., Sheehan, M., Clegg, C. W., & West, M. (2004, March 1). On the validity of subjective measures of company performance. *Personnel Psychology*, Vol.

57, pp. 95–118. https://doi.org/10.1111/j.1744-6570.2004.tb02485.x
Wang, C.-H., Moreau, D., & Kao, S.-C. (2019). From the lab to the field: Potential applications of dry EEG systems to understand the brain-behavior relationship in sports. *Frontiers in Neuroscience*, *13*(AUG), 893. https://doi.org/10.3389/fnins.2019.00893