2017

Testing of Wrist-Worn-Fitness-Tracking Devices During Cognitive Stress: A Validation Study

Nicole S. Chudy
University of Central Florida

TESTING OF WRIST-WORN-FITNESS-TRACKING DEVICES DURING COGNITIVE STRESS: A VALIDATION STUDY

by

NICOLE S. CHUDY

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Psychology in the
College of Sciences and in The Burnett Honors College
at the University of Central Florida
Orlando, Florida

Spring Term 2017

Thesis Chair: Dr. Daniel S. McConnell

ABSTRACT

The intent of this thesis was to test if one of the many fitness-tracking devices, Microsoft's Band 2 (MSB2), is accurate and reliable in detecting changes in Heart Rate (HR) and R-R intervals, during the repeated trial of two conditions of a working-memory test known as the N-Back. A 2 (devices: ECG, MSB2) × 4 (epochs: baseline 1, 1-back task, baseline 2, 3-back task) repeated measures factorial design was conducted. The participants were simultaneously equipped to the MSB2 and an electrocardiogram (ECG). The results of this study validated the MSB2 for the use in a cognitive task. The study suggests that fitness-tracking devices with similar sampling rates and features are candidates for further exploration as alternatives to ECG, in hope of making the inclusion of physiological data in psychological research more available and accessible.

DEDICATION

I'd like to dedicate this thesis to my mentor and committee thesis chair, Dr. Daniel S. McConnell. His guidance and belief in my abilities immensely contributed to the completion and success of this thesis, as well as opening my eyes to the vast domains of knowledge and the hard-work required to achieve greater learning. He truly is an inspiration and I greatly respect him.

I'd also like to dedicate this thesis to my beautiful family, my significant other David, my friend Juan, and all my other dear friends for their unconditional love and support. Through the periods of being lost in deep waters, they helped pull me up closer to the sun.

ACKNOWLEDGMENTS

I would like to acknowledge Dr. Janan A. Smither for allowing me to use the Technology and Aging Laboratory to conduct the experimentations, utilize her resources, and for her care and concern regarding the project's progress.

I would like to thank my committee members, Dr. Bedwell and Dr. Gordon, for their time and effort.

I would like to thank my brilliant colleagues & friends in the Technology and Aging Lab at UCF for their support and assistance.

I would like to thank the Office of Undergraduate Research for helping to sponsor this project, without their funding, this project would not have reached its potential.

TABLE OF CONTENTS

INTRODUCTION 1

LITERATURE REVIEW 3

METHOD 6

Participants 6

Materials & Apparatus 6

Procedures 7

RESULTS 9

HR 9

R-R 11

GSR 13

Task Performance 13

Questions on Task Difficulty 13

DISCUSSION 15

REFERENCES 17

LIST OF FIGURES

Figure 1: Descriptive statistics with 95% confidence intervals for HR..10

Figure 2: Descriptive statistics with 95% confidence intervals for RR intervals 12

LIST OF TABLES

Table 1: Concordance correlation coefficients with 95% confidence intervals for HR 4

Table 2: Descriptive statistics with 95% confidence intervals for HR 9

Table 3: Descriptive statistics with 95% confidence intervals for RR intervals............................11

INTRODUCTION

The demand for consumer-level fitness-tracking devices has increased in recent years. These devices, such as the Fitbit Charge HR, Polar H7, and Microsoft Band are strapped to the wrist or chest and are typically used for athletic and health purposes. Such devices claim to track variables such as heart rate (HR), R-R intervals (duration between peaks in the QRS complex ECG waveform), and galvanic skin response (GSR). HR is typically derived by counting the number of beats per minute (Engström, Ottosson, Wohlfart, Grundström, & Wisén, 2012). R-R intervals are collected by examining the amount of time in milliseconds (ms) between consecutive heart beats and are used as the basic of calculations of heart rate variability (HRV) across consecutive R-R intervals (Föhr et al., 2015). R-R intervals and HR data are accepted and objective physiological measures of stress and workload (Föhr et al., 2015; Laukkanen, & Virtanen, 1998). Higher HR, and consequently shorter R-R intervals are associated with higher arousal. Galvanic skin response (GSR) is also an established physiological measure of arousal (Shi, Ruiz, Taib, Choi, & Chen, 2007).

ECG is the gold standard for detecting physiological variables (Engström et al., 2012). ECG has limitations in many environments due to the size and cost of the equipment as well as the technical training needed to operate the machine. The researcher must be trained regarding the use of ECG and must be aware of the participants' movements during the experiment. Even simple movements can throw off the accuracy of the data.

ECG gathers cardiac information by detecting bioelectric signals from the heart through different locations on the body. Therefore, ECG machines are discomforting due to the attachment

of electrodes on multiple places on the body. This introduces an extraneous variable and potentially diminishes the validity of any obtained measures.

The Microsoft Band 2, along with other consumer-level fitness-tracking devices, use a different technique to analyze physiological variables that enables them to be small enough to be worn. This technique, called photoplethysmography (PPG) involves emitting light that is absorbed by the wearer's blood (Spierer, Rosen, Litman, & Fujii, 2015; Gregoski et al., 2012). The light absorption is greatest when there is maximum blood flow through the circulatory system after a heartbeat. The device can measure the light that bounces back to determine when maximum absorption occurs, allowing an estimation of when heart beats occur and how often.

With these differences in mind, this study seeks to validate the accuracy and reliability of fitness-tracking devices in measuring physiological data. A viable, cheaper alternative to ECG would make the inclusion of physiological data in psychological research more available and accessible.

LITERATURE REVIEW

Portable heart rate monitors prepared with conventional electrodes have been available since 1983 (Laukkanen & Virtanen, 1998). Past empirical research studies have demonstrated that certain fitness-tracking devices are reliable in terms of tracking and analyzing data (Goodie, Larkin, & Schauss, 2000; Wang et al., 2017). Multiple past empirical studies have measured HR by placing a belt with electrodes around the chest while a wrist worn unit records the data (Engström et al., 2012).

The first cardiac tracking devices started with HR made by Polar, which is one of the leading companies for HR monitors. One study assessed the Polar Vantage XL HR monitor in tracking physiological variables during a stationary task. Results indicated that the average HR between ECG and the Polar Vantage XL HR monitor were positively correlated (Goodie et al., 2000). However, the Polar device sampled HR at 0.20 Hz; modern devices can sample HR as fast as 1 Hz, which can provide a more temporally precise measure of HR. Another limitation of the Polar XL was that it was large and obtrusive when placed around the sternum; modern devices are instead conveniently strapped around the wrist.

In a sample of 50 healthy adults, the Apple watch, Mio fuse, Fitbit charge HR, and the Basis Peak were evaluated and compared to the Polar H7 (Wang et al., 2017). Participants were asked to walk on a treadmill at 2 mph, 3 mph, 4 mph, 5 mph, and 6 mph. A total of 1773 HR samples were recorded and results indicated that there was a significant positive correlation for all devices (Wang et al., 2017).

Table 1

Concordance correlation coefficients with 95% confidence intervals for HR from the study by Wang, et al., 2017.

	Agreement With Electrocardiogram
Device	Concordance Correlation Coefficients (95% CI)
Polar H7	.99 (.987-.991)
Apple Watch	.91 (.884-.929)
Mio Fuse	.91 (.882-.929)
Fitbit Charge HR	.84 (.791-.872)
Basis Peak	.83 (.779-.865)

While HR data is useful when studying cardiac output during physical activity, it is also useful in studying the arousal response during various cognitive tasks. Difficult cognitive tasks increase stress and arousal, including memory (Johnsen, Hansen, Murison, Eid, & Thayer, 2012). and attention/vigilance tasks (Luque-Casado, Zabala, Morales, Mateo-March, & Sanabria, 2013). Thus, obtaining physiological measures can be useful to experimental psychologists. The current question is whether consumer fitness devices are as effective as ECG in detecting such task-related changes in stress and arousal.

In addition to fitness-tracking devices, certain current generation smartphones also provide physiological information such as HR. This is done by using an application, placing a finger on the cell phone's camera, and then a sensor uses PPG to record HR. In a study involving Android phones, HR was recorded during three conditions involving both mental and physical stress (at rest, reading out-loud, and playing a video game) (Gregoski et al., 2012). Results indicated that the Android acquisition application was effective and consistent in detecting variations in HR across the three conditions, and was significantly correlated with ECG measures (Gregoski et al., 2012). This study suggests that fitness-tracking devices can be used during cognitive testing.

A study done at Cambridge University created a business cognitive application named Cognition Kit for the MSB2 and validated its use by using the N-Back task with two conditions (1-back, 2-back) (Cormack, Taptiklis, Barnett, King, & Fenhert, 2016). However, this study did not provide physiological results nor validated the MSB2 in terms of its physiological output. Also, the study used a sample size of only 10 participants and did not counterbalance the tasks for order effects (Cormack et al., 2016).

There thus exists a gap in the current literature regarding the effectiveness of using wrist-worn fitness-tracking devices in obtaining valid cardiac data during cognitive tasks. The goal of the current study is to address this by measuring heart rate activity during counterbalanced cognitive tasks using both ECG and fitness devices, and comparing the measurements between devices and across tasks.

METHOD

Participants

There were a total of 49 students from the University of Central Florida, between the ages of 18-32, recruited for this study using the SONA system. Participation was in exchange for partial or extra course credit. Thirty females (M_{age} = 18.67, *SD* = 1.69), and 19 males (M_{age} = 21.26, *SD* = 4.39), without visual deficiencies or on stimulant medication participated in this study. Participants were also required not to engage in vigorous exercise prior to the study, as well as avoid caffeinated beverages. All protocols were approved by the university IRB, and participants provided informed consent prior to participation.

Materials & Apparatus

Participants were seated at a workstation with a Dell E176FP 43.2cm flat panel LCD monitor with a refresh rate of 60Hz and a resolution of 1280x1024 pixels. The seating was standardized at approximately 65cm from the participant to the monitor.

N-Back Task: The N-back task is a working memory test that presents a series of letters and requires the participant to recall previous letters in the sequence (Kane, Conway, Miura, & Colflesh, 2007). The task requires the participant to keep the sequence in memory, and to respond when the current stimulus matches a particular previous stimulus. Task difficulty is determined by the number of intervening digits between the target stimulus and the to-be-remembered stimulus. The easy version of the task, referred to as the 1-back, asks participants to recall the previous stimulus while the more difficult 3-back requires participants to recall three previous stimuli. Participants indicate when a match occurs by clicking the left mouse button using their dominant hand. The task was presented using E-Prime software.

Microsoft Band 2 (MSB2): A medium sized band was used to acquire HR and RR interval data. Placement for the MSB2 was four centimeters down the participant's wrist on the hand that was non-dominant. An Android data acquisition application was developed using the MSB SDK, allowing extraction of GSR, HR and R-R interval data.

Basis Peak: This device was originally planned to be used in the study alongside the MSB2, but was abandoned due to a safety recall. Heart rate data was obtained using this device from a portion of the participants, but was not further analyzed.

Electrocardiogram (ECG): A Biopac Model MP150 device was used to gather HR data for each participant. HR data and HRV data was sampled at 500Hz. Physiological data was processed from ECG through the AcqKnowledge data acquisition system (BIOPAC Systems, Inc., Goleta, CA) using the built in noise filtering with frequency cut-offs, which was then transferred to the Kubios HRV software for analysis and to replicate 5-minute and 11-minute epochs for analysis. Due to faulty leads, GSR data was not obtained using the Biopac machine.

Questions on Task Difficulty: Participants were required to answer two questions on task difficulty following the N-back tasks. Difficulty was assessed on a 7-point scale, with 1 being considered easy and 7 considered difficult.

Card sorting: As a measure of basic cognitive ability (Falduto & Baron, 1986), participants were asked to sort a shuffled deck of 52 standard playing cards.

Procedures

Participants were first presented with the card sorting task. All participants passed this pre-requisite and were then able to continue with the study.

Participants were randomly assigned to either the (1-Back, 3-Back) condition or the counterbalanced (3-Back 1-Back) condition. Next, participants were simultaneously attached to the MSB2 and ECG devices.

The participants were then given a 5-minute baseline recording prior to the first task. An 11-minute N-back task was administered as the devices recorded heart rate data. Following this procedure, participants were then given a second baseline to recover from any task-induced stress and return to their physiological resting state. Next, participants performed the other N-back task for another 11-minute epoch. After each N-back task, participants were asked to judge the task difficulty. Participants were debriefed at the conclusion of the study session, which lasted approximate 60 minutes.

RESULTS

Statistical results were analyzed using Microsoft Excel and SPSS. We conducted two separate 2 (device: MSB2, ECG) × 4 (epochs: Baseline 1, 1-Back, Baseline 2, 3-Back) repeated measures ANOVAs to analyze the heart rate data. The dependent variables were HR and R-R intervals. Sphericity was assumed.

HR

Descriptive statistics for the associated HR data can be seen in Table 2 and Figure 1.

Table 2

Mean HR data as a function of task for each of the devices.

	Epoch			
Devices	Baseline 1	1-Back	Baseline 2	3-Back
ECG	*M* = 78.70	*M* = 79.88	*M* = 79.36	*M* = 80.26
	SE = 1.70	*SE* = 1.72	*SE* = 1.65	*SE* = 1.70
MSB2	*M* = 77.09	*M* = 78.55	*M* = 77.29	*M* = 78.76
	SE = 1.53	*SE* = 1.66	*SE* = 1.48	*SE* = 1.60

Note. M is the mean in beats/min. SE is standard error.

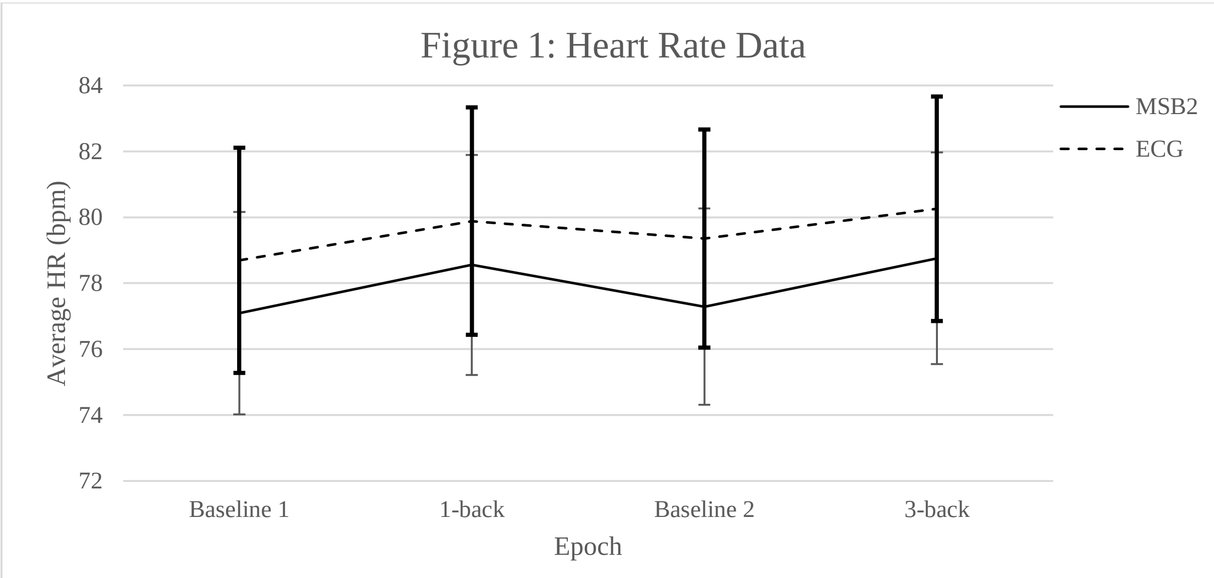

Figure 1: Heart rate data, per device, across the four epochs with 95% CI.

There was a significant main effect of epoch on HR, $F(3, 144) = 3.82$, $p = .011$, $n_p^2 = .07$, and a significant main effect of device, $F(1,48) = 42.07$, $p < .001$, $n_p^2 = .47$. In addition, there was evidence of an interaction between the devices and epoch, $F(3, 144) = 3.12$, $p = .028$, $n_p^2 = .06$. The interaction can be seen in Figure 1, in that the drop in HR for the second baseline was more pronounced as measured by the MSB2 than by the ECG. While this interaction suggests that the devices responded differently to the changing task, we note the effect size was negligible.

LSD post-hoc analyses showed that the marginal mean for HR on the 1-back task was significantly higher ($M = 79.22$, $SE = 1.69$) than that of the baseline 1 condition ($M = 77.89$, $SE = 1.61$), $M_d = 1.33$, $SE = 0.59$, $p = .029$. The 1-back task was not significantly different from the baseline 2 condition ($M = 78.32$, $SE = 1.56$), $M_d = 0.90$, $SE = 0.58$, $p = 1.29$. The marginal mean for HR on the 3-back task ($M = 79.51$, $SE = 1.64$) was significantly higher than that of the baseline 1 condition ($M_d = 1.62$, $SE = 0.54$, $p = .004$) and the baseline 2 condition ($M_d = 1.18$, $SE = 0.52$, *p*

= .027), but not statistically different from the 1-back task, M_d = 0.29, *SE* = 0.38, *p* = .446. The baseline 2 condition was not significantly different from baseline 1 condition, M_d = 0.43, *SE* = 0.64, *p* = .502.

Average HR displayed a strong positive correlation between the two devices on the 1-back task, $r(49) = .99$, $p < .001$. The adjusted coefficient of determination was .99, representing a large effect. Mean HR also showed a strong positive correlation between devices on the 3-back task, $r(49) = .99$, $p < .001$. The adjusted coefficient of determination was .99, indicating a large effect.

R-R Interval

Descriptive statistics for the associated R-R data can be seen in Table 3 and Figure 2.

Table 3

Mean R-R intervals as a function of task and device.

	Epoch			
Devices	Baseline 1	1-Back	Baseline 2	3-Back
ECG	*M* = 784.01	*M* = 770.66	*M* = 776.91	*M* = 766.36
	SE = 113.45	*SE* = 106.49	*SE* = 105.06	*SE* = 105.45
MSB2	*M* = 788.15	*M* = 775.14	*M* = 784.122	*M* = 773.12
	SE = 106.34	*SE* = 104.80	*SE* = 97.89	*SE* = 102.13

Note. M is the mean in milliseconds. SE is the standard error.

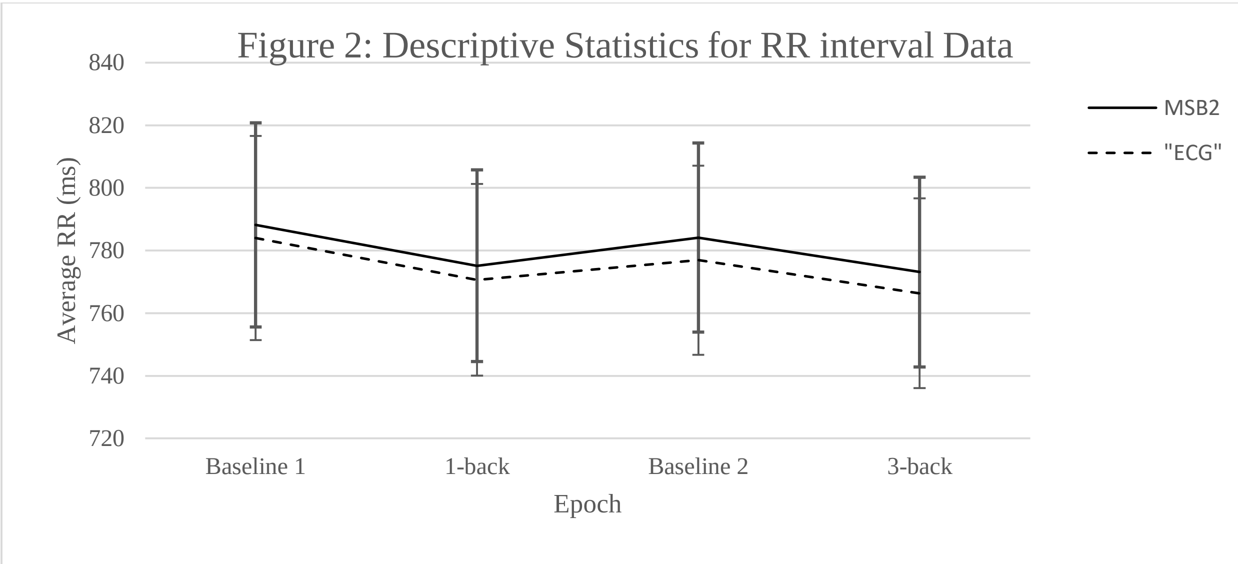

Figure 2: R-R intervals as a function of task and device with 95% CI.

Results obtained from R-R measurements gave evidence of a significant main effect for times of measurement, $F(3, 144) = 3.44$, $p = .019$, $n_p^2 = .07$, and a significant main effect for devices $F(1,48) = 10.31$, $p = .002$, $n_p^2 = .18$. Results indicated that there was no significant interaction between the devices and times of measurement in detecting task differences, $F(3, 144) = 0.85$, $p = 4.71$ $n_p^2 = .02$.

LSD post-hoc analyses showed that the marginal mean for R-R intervals on the 1-back task was significantly lower ($M = 772.90$, $SE = 15.07$) than the baseline 1 condition ($M = 786.08$, $SE = 15.65$), $M_d = -13.18$, $SE = 6.27$, $p = .041$. The 1-back task was not significantly different from baseline 2 condition ($M = 780.52$, $SE = 14.45$), $M_d = -7.62$, $SE = 5.90$, $p = .203$ The marginal mean for R-R intervals on the 3-back task ($M = 769.74$, $SE = 14.80$) was significantly lower than the baseline 1 condition ($M_d = -16.34$. $SE = 5.60$, $p = .005$), and baseline 2 condition ($M_d = -10.78$, $SE = 4.73$, $p = .027$), but was not significantly different from the 1-back task, $M_d =$

-3.16, $SE = 3.64$, $p = 3.89$. There was no significant difference between the baseline 2 condition and baseline 1 condition, $M_d = 5.57$, $SE = 7.05$, $p = 4.34$.

Mean R-R established that there was a strong positive correlation between the devices on the 1-back task, $r(49) = .995$, $p < .001$. The adjusted coefficient of determination was .995. signifying a large effect. Mean R-R data also showed strong positive correlation between the devices on the 3-back task, $r(49) = .995$, $p < .001$.The adjusted coefficient of determination was .995, representing a large effect.

GSR

A one-way ANOVA was conducted on the GSR data from the MSB2. Sphericity was assumed. Results indicated that there was no significant main effect for times of measurement, $F(3, 144) = 1.18$, $p = .320$, $n_p^2 = .02$.

Task Performance

Performance on the N-back was analyzed using Signal Detection Theory. Correct detections of a stimulus match were considered hits, while positive responses that did not correspond to a match were considered false alarms. Sensitivity (d') was calculated as the hit rate minus the false alarm rate. A paired-samples t-test was conducted to compare d' between the 1-back and 3-back tasks. There was a significant difference in the scores for the 1-back ($M = 3.90$, $SE = 0.12$) and the 3-back ($M = 2.08$, $SE = 0.16$) conditions. Results suggest that participants performed significantly better in the 1-back condition than the 3-back, $t(48) = 11.33$, $p < .001$, $d = 1.86$.

Questions on Task Difficulty

A paired-samples t-test was conducted to compare the participants' perceptions on task difficulty. Results demonstrated a significant difference between scores on the 3-back task (M = 5.31, SE = 0.14) and the 1-Back task (M = 1.59, SE = 0.10). Results suggest that participants subjectively perceived the 3-back task as more difficult compared to the 1-back task, $t(48) = 23.65$, $p < .001$, $d = 4.41$, indicating a large effect.

DISCUSSION

The MSB2 records HR at 1Hz, GSR at 0.2/5 Hz, and R-R intervals at varied intervals. The ECG we used in the study records data up to 500 Hz. According to the Nyquist-Shannon Theorem, in order to get an accurate reading and representation of information outputted from a digital-analog device, the sampling rate must be twice that of the maximum frequency (Lindner, 2009; Proakis & Manolakis 2007). The MSB2 obtaining a frequency of 1 Hz may account for the interaction we observed between the devices and times of measurements for HR data, but none for R-R intervals. This suggests that the devices did not detect differences in the exact same way for HR data. The more pronounced drop in HR for the second baseline, as measured by MSB2, may have been an artefact, though we repeat that the effect size of the interaction was negligible. While there was a slight difference in average HR measured between the devices, on the whole, there was a high degree of concordance between the ECG and MSB2 measures of HR and R-R intervals across the four epochs recorded in the study. This indicates that the MSB2 was as effective as the ECG in measuring task-induced variations in heart rate.

A potential issue, however, is the very small degree of task-induced variations. The GSR, HR, and R-R interval data failed to register significant differences across the two N-back tasks. Variations only occurred when switching between the task and the resting periods. Even in this case, the differences were only significant for HR and R-R interval data, and the effect size was very small. These results are not consistent with previous studies that measured larger changes in physiological measures of stress in response to variations in the difficulty of the N-back task (Mandrick, Peysakhovich, Rémy, Lepron, & Causse, 2016). Despite the lack of physiological

findings, participants significantly perceived the 3-back task as more cognitively demanding and stressful in comparison to the 3-back task.

Implications for future research involve further physiological testing for the N-Back task, among other working memory tests. Perhaps a more difficult task must be administered to induce variations in physiological arousal. A lack of incentive urgency or consequence may also account for the absence in arousal difference between some of our epochs. Further, future work can determine whether the device can detect task-induced variations in heart rate when participants are moving, which is a current limitation of ECG technology.

Limitations regarding the MSB2 include being limited to the medium sized band (there are multiple sizes due to differences in wrist size), random inconsistencies with device-to-cellular synchronization, and less photosensitive skin (Spierer, Rosen, Litman, & Fujii, 2015). A large sample size was chosen to minimize these errors. It is worth noting that in addition to the safety recall that lead us to abandon use of the Basis Peak device, the MSB2 has been discontinued by Microsoft at the time of writing. Experimenters wishing to use such devices for research purposes may require continued device support in order to use them long-term – however, when devices are discontinued, manufacturer support may be lost. While our results suggest that the devices may be valid for use in the psychological laboratory, their availability and support may be subject to market forces that limit their long-term viability as research tool.

REFERENCES

Cormack, F., Taptiklis, N., Curtis, J., van Schaik, T., Glazer, E., Fehnert, B., King, J., & Barnett, J. (2016). High-Frequency Monitoring of Cognition, Mood and Behaviour Using Commercially Available Wearable Devices. *age*, *30*(40), 50.

Engström, E., Ottosson, E., Wohlfart, B., Grundström, N., & Wisén, A. (2012). Comparison of heart rate measured by Polar RS400 and `ECG, validity and repeatability. *Advances in Physiotherapy, 14(3),* 115-122.

Falduto, L. L., & Baron, A. (1986). Age-related effects of practice and task complexity on card sorting. *Journal of Gerontology, 41(5),* 659-661.

Föhr, T., Tolvanen, A., Myllymäki, T., Järvelä-Reijonen, E., Rantala, S., Korpela, R., Peuhkuri, K., Kolehmainen, M., Puttonen, S., Lappalainen, R. and Rusko, H. (2015). Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. *Journal of Occupational Medicine and Toxicology*, *10*(1), 39.

Goodie, J. L., Larkin, K. T., & Schauss, S. (2000). Validation of the Polar Heart Rate Monitor for Assessing Heart Rate During Physical and Mental Stress. *Journal of Psychophysiology, 14(3),* 159-164.

Gregoski, M.J., Mueller, M., Vertegel, A., Shaporev, A., Jackson, B.B., Frenzel, R.M., Sprehn, S.M. and Treiber, F.A. (2012). Development and validation of a smartphone heart rate acquisition application for health promotion and wellness telehealth applications. International journal of telemedicine and applications, 2012, 1.

Johnsen, B. H., Hansen, A. L., Murison, R., Eid, J., & Thayer, J. F. (2012). Heart rate

variability and cortisol responses during attentional and working memory tasks in naval cadets. *Int Marit Health*, *63*(4), 181-187.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (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.

Laukkanen, R. M., & Virtanen, P. K. (1998). Heart rate monitors: State of the art. *Journal of Sports Sciences, 16*(1), 3-7.

Lindner, D. (2009). *Introduction to Signals and Systems*. New York City, NY: McGraw Hill Company.

Luque-Casado, A., Zabala, M., Morales, E., Mateo-March, M., & Sanabria, D. (2013). Cognitive Performance and Heart Rate Variability: The Influence of Fitness Level. *PLoS ONE, 8*(2), e56935.

Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology, 121,* 62-73.

Proakis, J., & Manolakis, D. (2007). *Digital signal processing: principles, algorithms, and Applications*. Upper Saddle River, NJ: Prentice-Hall.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems* (pp. 2651-2656). ACM.

Spierer, D., Rosen, Z., Litman L., & Fujii, K. (2015). Validation of photoplethysmography as a

method to detect heart rate during rest and exercise. *Journal of Medical Engineering & Technology*, 39(5), 264-271.

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2016). Accuracy of Wrist-Worn Heart Rate Monitors. *Jama cardiology*.