Validation of photoplethysmography using a mobile phone application for the assessment of heart rate variability in the context of heart rate variability-biofeedback

van Dijk, W.; Huizink, A.C.; Oosterman, M.; Lemmers-Jansen, I.L.J.; de Vente, W.

DOI
10.1097/PSY.0000000000001236

Publication date
2023

Document Version
Final published version

Published in
Psychosomatic Medicine

Validation of Photoplethysmography Using a Mobile Phone Application for the Assessment of Heart Rate Variability in the Context of Heart Rate Variability–Biofeedback

Willeke van Dijk, MSc, Anja C. Huizink, PhD, Mirjam Oosterman, PhD, Imke L.J. Lemmers-Jansen, PhD, and Wieke de Vente, PhD

ABSTRACT

Objective: Heart rate variability–biofeedback (HRV-BF) is an effective intervention to reduce stress and anxiety and requires accurate measures of real-time HRV. HRV can be measured through photoplethysmography (PPG) using the camera of a mobile phone. No studies have directly compared HRV-BF supported through PPG against classical electrocardiogram (ECG). The current study aimed to validate PPG HRV measurements during HRV-BF against ECG.

Methods: Fifty-seven healthy participants (70% women) with a mean (standard deviation) age of 26.70 (9.86) years received HRV-BF in the laboratory. Participants filled out questionnaires and performed five times a 5-minute diaphragmatic breathing exercise at different paces (range, ~6.5 to ~4.5 breaths/min). Four HRV indices obtained through PPG, using the Happitech software development kit, and ECG, using the validated NeXus apparatus, were calculated and compared: RMSSD, pNN50, LFpower, and HFpower. Resonance frequency (i.e., optimal breathing pace) was also compared between methods.

Results: All intraclass correlation coefficient values of the five different breathing paces were "near perfect" (>0.90) for all HRV indices: lnRMSSD, lnpNN50, lnLFpower, and lnHFpower. All Bland-Altman analyses (with just three incidental exceptions) showed good interchangeability of PPG- and ECG-derived HRV indices. No systematic evidence for proportional bias was found for any of the HRV indices. In addition, correspondence in resonance frequency detection was good with 76.6% agreement between PPG and ECG.

Conclusions: PPG is a potentially reliable and valid method for the assessment of HRV. PPG is a promising replacement of ECG assessment to measure resonance frequency during HRV-BF.

Key words: photoplethysmography, diaphragmatic breathing, heart rate variability, mobile app, validation, electrocardiography.

INTRODUCTION

Heart rate variability–biofeedback (HRV-BF) is a promising method to manage physiological stress and promote mental health (1,2). The idea behind HRV-BF is that the practice of slow diaphragmatic breathing while receiving feedback on beat-to-beat heart rate (HR) can help to increase respiratory sinus arrhythmia (RSA), which is the vagally mediated increase of HR during inhalation and decrease during exhalation. RSA is one of the contributors to heart rate variability (HRV), defined as the variations in beat-to-beat HR measured by consecutive R-wave peaks (3). HRV is a commonly used marker of mental and physiological stress and is a relevant outcome for mental health research (4,5). Higher HRV indicates increased adaptive capacity or resilience, as it demonstrates sensitivity to the environment and the ability to maintain homeostasis (6,7). In contrast, low HRV may imply an imbalance between the parasympathetic and sympathetic nervous systems and has been associated with increased or chronic stress and cardiovascular risks (8–10).

ECG = electrocardiography, **HFms^2** = high-frequency power in milliseconds squared, **HRV** = heart rate variability, **HRV-BF** = heart rate variability–biofeedback, **ICC** = intraclass correlation coefficient, **LFms^2** = low-frequency power in milliseconds squared, **PPG** = photoplethysmography, **RMSSD** = square root of mean squared differences of successive normal-to-normal intervals, **pNN50** = percentage of successive normal sinus beat to beat intervals more than 50 milliseconds, **SDK** = Software Development Kit

In addition to RSA, HRV is affected by the baroreflex, which is a reflex that modulates blood pressure through slightly increasing and decreasing HR. As explained by Lehrer et al. (2), the baroreflex contributes to blood pressure homeostasis and is, like RSA,

SDC Supplemental Digital Content

From the Departments of Clinical, Neuro and Developmental Psychology (van Dijk, Huizink, Lemmers-Jansen) and Clinical Child and Family Studies (Oosterman), Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam; Institute for Brain and Behavior Amsterdam (IBBA), Amsterdam, the Netherlands; Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom (Lemmers-Jansen); and Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, the Netherlands (de Vente).

Address correspondence to Willeke van Dijk, MSc, Room MF-A529, Van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands. E-mail: w.van.dijk@vu.nl

Received for publication September 1, 2022; revision received May 25, 2023.

Article Editor: Mirela Habibović (Guest Editor)

DOI: 10.1097/PSY.0000000000001236

under parasympathetic control. Because of a time delay in the baroreflex of ~5 seconds between sensing deviant blood pressure in the arteries and a cardiac response, a periodic rhythm of increasing and decreasing HR occurs lasting ~10 seconds. This rhythm varies among individuals between ~4.5 and ~6.5 cycles/min (e.g. (11)). Breathing at the same pace as the baroreflex system's resonance frequency, a biophysical characteristic of any "closed-loop control system with feedback" (11), creates a large variation in HR, which stimulates the baroreflex. This breathing at resonance frequency in short is thought to increase baroreflex efficiency, which in turn improves autonomic activity modulation (12). Accordingly, breathing at resonance frequency is regarded as the optimal pace to obtain beneficial health effects, and breathing exercises at resonance frequency is therefore the core of HRV-BF interventions.

Before HRV-BF intervention, an individual's resonance frequency needs to be identified, which is mostly done through spectral analysis of heart periods (i.e., the time in milliseconds between subsequent heart beats, reflecting beat-to-beat HR; (12)). The rhythmicity of the baroreflex between ~4.5 and ~6.5 cycles/min converges largely with the low-frequency (LF) spectral power band of heart periods of 0.04 to 0.15 Hz (~2.5–9 cycles/min; (13)). Hence, to detect resonance frequency, an individual's breathing pace that results in the largest amplitude (peak) in the LF band needs to be identified (12).

To provide feedback on HRV and detect resonance frequency, HRV-BF is often offered through technologies that require an elaborate electrocardiogram (ECG) setup, measuring the electrical signal associated with each heartbeat. However, because of the substantial costs and lack of user-friendliness, this method hampers the accessibility of HRV-BF by a broad public in daily life. Recent technological advances have increased the accessibility of HRV-BF as more and more mobile devices allow individuals to monitor their health, HR, and HRV (see, for a review, Ref. (14)). These devices usually measure cardiac activity using an infrared camera on the wrist (smartwatch) or fingertip (apps using the camera of the mobile phone). Measuring cardiac activity using an infrared camera is called photoplethysmography (PPG) (15). Through PPG, beat-to-beat blood volume changes in the microvasculature of peripheral tissues can be detected, which can be used to calculate HR and HRV. A huge advantage of this technique is that it is cheap, easy to apply, wireless, and portable.

Despite the clear advantages of applying HRV-BF using PPG-derived HRV, there is concern about the ability of these wearable devices to correctly determine clinical biomarkers such as HR and HRV. Moreover, to date, none of the existing apps that claim to measure HR and HRV have been marked yet with CE (Conformité Européenne) classification, which indicates that the product meets the safety, health, and environmental requirements appointed by the European Union. Accurate and detailed measurements of heart periods are essential to determine HRV and support high-quality HRV-BF interventions. To date, several studies have tested the validity of PPG-based HRV in comparison to the "gold standard" ECG assessments during a variety of procedures (e.g., rest, walking, attentional load) and positions (e.g., standing, seated, supine) and showed high degree of agreement between the methods (e.g., (16–23)). However, no studies have compared PPG and ECG specifically during slow diaphragmatic breathing, which is the key technique that is taught during HRV-biofeedback interventions.

Therefore, in this study, we aimed to validate the use of PPG to measure cardiac activity in healthy individuals by comparing PPG assessments with simultaneously measured ECG. The current study compares a novel Software Development Kit (SDK), developed by Happitech (Amsterdam, the Netherlands), for PPG measurement through a smartphone camera, with simultaneous ECG measurements. This SDK is the first CE-certified smartphone application for HRV measurements through PPG. The SDK is implemented in an application that supports women to quit smoking through stress reduction, which is currently under investigation (see van Dijk et al. (24) for the study protocol). If PPG seems to be a valid method to measure HRV during slow, diaphragmatic breathing, this can promote research into the effectiveness and implementation of HRV-BF interventions offered through mobile apps using PPG. If proven reliable and effective, PPG-based HRV-BF interventions contribute to the availability of low-cost, accessible stress-reducing interventions to the broad public.

In sum, the research question of this study was whether HRV can be reliably and validly measured through PPG with use of a smartphone camera during slow, diaphragmatic breathing. Reliability was determined through comparison of HRV indices and identification of resonance frequency between ECG- and PPG-based heart periods. The ECG signal was processed using NeXus apparatus (NeXus-4; Mind Media, Herten, the Netherlands), which is also used for HRV-BF purposes, and the PPG signal was processed using a recently developed SDK by Happitech (25).

We expected that PPG-HRV and ECG-HRV indices would not differ significantly from one another and therefore would strongly correlate. Moreover, we expected that PPG- and ECG-based identification of resonance frequency would highly correspond within individuals. The results of Kim et al. (26) that demonstrate significant increases in the LF peak from pre– to post–HRV-BF training imply that experience in diaphragmatic breathing may facilitate resonance frequency detection. Based on these findings, we expected to find a higher resonance frequency correspondence in participants with, compared to without, experience in diaphragmatic breathing obtained through, for example, yoga. Because subject characteristics such as age, sex, weight, and height have been shown to influence HRV indices and resonance frequency (11,27), we also investigated whether these variables affected correspondence in resonance frequency.

METHODS

Participants and Recruitment

Participants were either acquaintances of the researchers or Bachelor Psychology students, who signed up to participate to earn research participation points. Sixty-six healthy Dutch speaking participants (50% women) were recruited, aged between 18 and 60 years. Data of nine participants were removed from the analyses. Data from five participants were unusable because of erroneous export of ECG data; from two participants, PPG data were lost because of problems with Internet connection; and from the remaining two participants, quality of PPG and ECG data was insufficient. The reason behind exclusion was that when data of either PPG and/or ECG are missing, comparison between PPG and ECG is no longer possible. The final sample consisted of 57 participants (mean age = 26.7 years; n = 40 [70%] women). Excluded participants from which background data could be retrieved did

not differ in age (mean [standard deviation], or M [SD] = 23.0 [4.73] years) from included participants (M [SD] = 25.9 [9.33] years), $t(61) = 0.79$, $p = .43$). Also, the distribution of sex was not different between the excluded and included groups (χ^2 (1, $N = 63$) = .45, $p = .43$). Thirty participants reported having previous experience with diaphragmatic breathing practices.

Trial Design

This study has a cross-sectional design. HRV measurements obtained through PPG using a smartphone camera were compared with simultaneously obtained HRV data by ECG using a NeXus apparatus. Power analyses using G*Power (28) revealed that a sample size of at least $N = 52$ is required to be able to detect small differences, defined according to common practice in biomedical research (Cohen, 1988 (29); Hopkins, 2009 (30); Cohen d within-group effect size = 0.4 with $\alpha = .05$ and $1 - \beta = 0.80$), between PPG- and ECG-derived HRV indices. Participants were tested during two different waves: wave 1 in November 2020 ($n = 10$) and wave 2 in December 2021 until February 2022 ($n = 56$). HRV indices measured through PPG and ECG were obtained during five different paced breathing exercises. Demographics and background information were measured using questionnaires that were administered before the breathing exercises. To validate the test procedure of slow diaphragmatic breathing exercises, positive and negative moods were measured before and after the breathing exercises using rating scales. Because slow-breathing exercises are known to induce immediate relaxation (e.g., (31–33)), observing reduction of tension and anxiety, as measured in the negative mood scale, and reduction of alertness and excitement, as measured in the positive mood scale, would indicate proper execution of the slow-breathing technique.

Written informed consent was obtained from all participants. All personal data were stored and protected according to the general data protection regulation guidelines. Participants were assigned identification numbers that were used throughout the study. Ethics approval was obtained from the "Vaste Commissie Wetenschap en Ethiek of the Vrije Universiteit van Amsterdam" (reference number VCWE-2019-161). Data are available upon request via the corresponding author.

Procedure

At the beginning of the 1-hour lasting experiment, participants were seated and informed about the study procedure. They were instructed by a research assistant to place three ECG electrodes themselves (one on each side right below the collarbone and one below the left rib cage), as a close contact with the participants was avoided, because of COVID-19 safety guidelines. The electrodes were then connected to the NeXus device, which was connected with the NeXus laptop via Bluetooth, and participants were asked to open the first exercise of the SDK application on the smartphone. Hereafter, participants filled out a questionnaire in the smartphone application, including demographic and background questions and questions about their mood. After completion, participants were notified that, during the following 30 minutes, they would remain seated and perform breathing exercises at six different paces for 5 minutes each, guided by a breath-pacer on the laptop screen placed in front of them. Because the first breathing pace was not part of the Lehrer protocol to detect resonance frequency (12), data from this breathing exercise are not reported in the present study. The five breathing exercises used for the present study consisted of the following paces (in breaths per minute): a) ~6.5 (inhalation 4 seconds, exhalation 5 seconds), b) 6.0 (inhalation 4 seconds, exhalation 6 seconds), c) ~5.5 (inhalation 4 seconds, exhalation 7 seconds), d) 5.0 (inhalation 4 seconds, exhalation 8 seconds), and e) ~4.5 (inhalation 5 seconds, exhalation 8 seconds).

The breath-pacer was depicted by a wave with an ascending line indicating inhalation and a descending line indicating exhalation. Participants were instructed to inhale through the nose and exhale through the mouth during the breathing exercises. They were encouraged to hold one hand on their belly to ensure diaphragmatic breathing; to sit upright, with both feet on the ground; and to relax their neck and shoulders during the breathing exercises. Throughout the procedure, the participants were discretely observed by one of the researchers to make sure the instructions were followed. After each breathing exercise, participants were asked to fill out questions in the app about their experience during the exercise and whether they felt nauseous or dizzy. Also, the research assistant asked how participants felt to ensure they were feeling comfortable throughout the session. Before they continued to the next exercise, participants were given the opportunity to pause and take something to drink for a maximum of 5 minutes to make sure their breathing would return to their normal pace.

After finishing all six breathing exercises, participants were again asked about their mood in the app. Subsequently, they were instructed how to disconnect the ECG electrodes and were thanked for their participation.

Instruments and Measures

PPG Recording

HRV data through PPG were obtained using the SDK developed by Happitech. This SDK uses the sensor of the rear camera of an enabled smartphone (iPhone 8 or Huawei P30 Lite), which assesses HRV data indices via the index or middle finger of the participant. Small changes in light intensity due to the amount of oxygen in the blood in the participant's tissue are detected, which is called pulse oximetry. These changes result in a waveform from which HRV measures can be calculated. The PPG measurement in the app was automatically initiated when participants correctly placed their finger on the camera.

ECG Recording

To acquire HRV data through ECG, we used the NeXus Hardware and BioTrace+ software (BioTrace+ for NeXus-4, version 2013; Mind Media, Herten, the Netherlands). The NeXus is an integrated system for HRV-BF and psychophysiological research, providing ECG data for both time and frequency domains, using self-adhesive electrodes. The researcher manually started the NeXus measurement when PPG measurement was initiated so that PPG and ECG data were collected simultaneously.

Background Characteristics

The following background characteristics were measured through self-report using a questionnaire: age, sex, highest obtained educational level (middle-level vocational training, higher vocational training, academic education, other), weight (in kilograms), and height (in centimeters). Background variables were dichotomized

into "high" and "low" groups using the nearest round value to the median: age of 25 years, weight of 65 kg, and height of 175 cm. In addition, a question concerning experience with diaphragmatic breathing through for example yoga or singing education was included (yes or no).

Mood

Mood was measured using 13 items derived from the Positive and Negative Affect Schedule (34), which measures positive and negative affectivity. To test the relaxing and calming effect that is expected to be induced by HRV-BF and/or slow breathing (35,36), we selected items that were characterized by positive and negative affect and by elevated levels of arousal (10 items for positive affect and 10 items for negative affect). Participants were instructed to indicate the extent to which the mood state correctly described their current feelings. Because HRV-BF may affect positive and negative valenced arousal to a different extent, we chose to calculate separate scores for positive and negative affect. The positive affect subscale consisted of the following six items: "excited," "alert," "determined," "active," "attentive," and "enthusiastic," with high scores indicating higher arousal and stronger positive affect. The negative affect subscale included the following seven items: "afraid," "upset," "distressed," "nervous," "jittery," "irritable," and "scared," with high scores indicating higher arousal and stronger negative affect. Internal consistency of both subscales was high, with Cronbach α values of .85 and .77 for positive and negative affectivity, respectively. To increase potential variability, items in the present study were scored on an 11-point Likert scale, ranging from "not at all" (0) to "very much" (10), rather than on 5-point Likert scales, as used in the original Positive and Negative Affect Schedule.

HRV Indices

Using both PPG and ECG instruments, we assessed two time-domain HRV indices: a) the square root of mean squared differences of successive normal-to-normal (N-N) intervals (RMSSD) and b) the proportion derived by dividing the NN50 by the total number of NN intervals (pNN50), and two frequency-domain indices: LFpower and HFpower. All four indices can be assessed within a 5-minute period (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996).

RMSSD

RMSSD is the beat-to-beat variance in HR in milliseconds (27). It is the principal measure within the time domain, used to estimate vagally mediated changes in HRV (37). Higher values indicate larger HRV.

pNN50

pNN50 represents the percentage of consecutive beat-to-beat (RR) intervals that vary by more than 50 ms (pNN50). It indicates the contribution of the parasympathetic nervous system to cardiac regulation (27). In people with normal sinus rhythm, a larger value of pNN50 indicates higher HRV.

LFpower

Low-frequency power was defined as the absolute power in the low-frequency band (0.04–0.15 Hz) in ms^2 ($LFms^2$) (27). LFpower reflects fluctuations in interbeat intervals (IBIs) ranging from ~7 to 25 seconds. The breathing frequencies that are involved in HRV-BF, which vary between 9 and 13 s/cycle, fall within this range.

HFpower

High-frequency power was defined as the absolute power in the high-frequency band (0.15–0.4 Hz) in ms^2 ($HFms^2$) (27). HFpower reflects fluctuations in IBIs ranging from 2.5 to ~7 seconds.

Resonance Frequency

Resonance frequency was defined as the breathing pace characterized by the highest power in the LF band, relative to the power in the total frequency spectrum (limited to 0.02–0.5 Hz, given the measurement duration). Because the main goal of HRV-BF is to train individuals to breathe at their resonance frequency, the breathing pace with the highest LF peak was identified and compared between PPG and ECG. Three variables were created. First, LF peak pace-PPG and LF peak pace-ECG reflected the breathing paces (range, 1–5) resulting in the highest LF peak percentage according to PPG or ECG assessment, respectively. In addition, a dichotomous variable was created reflecting congruence (1) versus incongruence (0) of resonance frequency between PPG and ECG assessment. As an individual's resonance frequency usually varies on a continuous scale between ~6.5 to ~4.5 breaths/min (11,38), LF peak percentages of two neighboring breathing paces are likely to be nearly identical. Hence, resonance frequency was considered congruent between PPG and ECG if identical breathing paces or directly neighboring breathing paces (e.g., pace 1 and 2) were identified as resonance frequency.

Extraction of PPG and ECG Signals

Main Calculation Flow for PPG and ECG Signals

For PPG, raw input (i.e., pixels) was extracted from the camera (frame rate = 30 fps). The data were processed by the following steps: pixel selection, interpolation (30 Hz > 180 Hz), and filtration using the Butterworth filter (pass band = [0.5–4] Hz; see Figure S1, Supplemental Digital Content, http://links.lww.com/PSYMED/A954). Then, R-peak detection of the filtered data was performed using trained neural networks (25), from which IBI information was extracted. For the calculation of IBIs from ECG data, a standard peak detection algorithm from MATLAB was used together with manual correction of erroneous calculated ECG peak position and marking and of signal segments of low quality due to artifacts (e.g., finger movement). Manual correction and marking were conducted using an application with a graphical user interface. The 5-minute intervals of both PPG and ECG data were exactly superposed (see Figure S2, Supplemental Digital Content, http://links.lww.com/PSYMED/A954). For data cleaning, erroneous/extreme IBIs due to incorrect identification of R peaks were excluded, after which an IBI correction algorithm was carried out, similar to data preprocessing as described by Widjaja et al. (39). Subsequently, frequency and time HRV indices were calculated based on cleaned and corrected IBIs according to standard formulas (40). For frequency analysis of HRV, the fast Fourier transformation spectrum was calculated using a Welch periodogram method (length of segment 5 minutes and 50% overlapping). Frequency bands were defined following the Task Force (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996).

Statistical Analysis

Statistical analyses were performed using IBM SPSS version 27. First, data were inspected for errors and outliers were checked by means of histograms. Erroneous data were removed, but outliers were kept in the data, following the guidelines described by Bland and Altman (41). The levels of agreement between RMSSD, pNN50, LFms^2, and HFms^2 from PPG and ECG were assessed using Bland-Altman analysis for each breathing exercise separately, and constant errors (CEs) were calculated. Similar to Charlot et al. (42), we decided a priori that acceptable limits of agreement for interchangeability between ECG and PPG corresponded to 150% of the reference SD values of ECG-derived RMSSD, LFpower, and HFpower, as provided by the Task Force (1996). Reliability of the PPG measurements was established by comparing them with the ECG measurements, using intraclass correlation coefficients (ICCs), separately for each HRV index, per both breathing pace (intrabreathing pace reliability) and overall breathing paces (interbreathing pace reliability). ICC values from 0 to 0.30 were classified as small, values from 0.31 to 0.49 as moderate, values from 0.50 to 0.69 as large, ICCs from 0.70 to 0.89 as very large, and values from 0.90 to 1.00 as near perfect (43). A significance level of $p < .05$ was used.

Correspondence between resonance frequency was determined using the ICC and through determining absolute agreement. To assess whether resonance frequency correspondence was dependent on the participants experience in diaphragmatic breathing, age, weight, and height, we performed χ^2 tests and reported the relative risk ("chance") to obtain resonance frequency among participants with and without diaphragmatic breathing experience, between women and men, and between groups "higher" and "lower" on age, weight, and height.

RESULTS

Initial Analyses

Information on the mood scales was missing for three participants, yielding data of a total of 54 participants on the mood scales. Visual inspection revealed that all variables' frequency distributions were approximately normal.

In accordance with the supposed relaxing and calming effect of slow diaphragmatic breathing, positive and negative valenced arousal decreased over the course of the experiment, giving some support for the validity of the test procedure (positive mood scale: M [SD]: pretest = 6.58 (1.42), posttest = 5.87 (1.49), $t(53) = 4.32$, $p < .001$; negative mood scale: M [SD]: pretest = 1.78 (1.27), posttest = 1.13 (0.94), $t(53) = 5.25$, $p < .001$).

Validation of PPG

Because of problems with either the PPG or ECG signal, 11 separate observations, consisting of data of a particular breathing pace from 10 participants, were excluded from the analyses. The means and SDs of RMSSD, pNN50, LFms^2, and HFms^2 of both the PPG and ECG measurements, and ICCs are displayed in Table 1. Full-validity statistics are shown in Tables S2–S5, Supplemental Digital Content, http://links.lww.com/PSYMED/A954. Because the distributions of RMSSD, pNN50, LFpower, and HFpower were slightly skewed to the right, the data were first ln-transformed before ICCs were calculated, for which the normality assumption holds. ICC values of the different breathing paces were "near perfect" (>0.90) for lnRMSSD, lnpNN50, lnLFms^2, and lnHFms^2. Bland-Altman plots per breathing pace per HRV index are presented in Figures 1A–D. Because the difference scores (of the nontransformed data) were normally distributed, a prerequisite for conducting Bland-Altman analyses (see, for example, Ref. (44)), further Bland-Altman analyses were conducted using nontransformed data. The CE and SEE values were consistent across breathing exercises for RMSSD and pNN50, with PPG-derived RMSSD and pNN50 consistently being slightly higher (i.e., about 3.5–5.5 milliseconds for RMSSD and 3% for pNN50) than ECG-derived RMSSD and pNN50, respectively (see Tables S2 and S3 [Supplemental Digital Content, http://links.lww.com/PSYMED/A954] and Figures 1A, B). The CE and SEE values for LFms^2 were small and nonsignificant, suggesting absence of systematic overestimation or underestimation of PPG (see Table S5 [Supplemental Digital Content, http://links.lww.com/PSYMED/A954] and Figure 1C). The CE and SEE values for HFms^2 were small to moderate and consistent over breathing paces, with PPG slightly overestimating HFms^2 (i.e., 3.5–7 ms^2; Table S5 and Figure 1D).

TABLE 1. Correspondence of PPG and ECG Measurements Per Parameter at Different Breathing Rates

			RMSSD		pNN50		LFms^2		HFms^2	
Breathing Pace		*N*	M (SD)	ICC	M (SD)	ICC	M (SD)	ICC	M (SD)	ICC
1: 4–5	PPG	53	59.53 (28.88)	0.97***	33.14 (19.11)	0.99***	265.41 (144.17)	0.99***	22.14 (18.60)	0.92 ***
	ECG		54.33 (28.01)		29.74 (19.72)		255.46 (141.88)		17.39 (15.59)	
2: 4–6	PPG	57	58.50 (25.73)	0.98***	32.52 (18.58)	0.96***	305.19 (153.92)	0.97***	25.94 (17.51)	0.86***
	ECG		54.73 (22.33)		29.62 (18.96)		292.39 (150.30)		19.36 (16.14)	
3: 4–7	PPG	55	56.89 (23.05)	0.98***	32.64 (19.38)	0.98***	338.23 (175.68)	0.98***	27.68 (16.26)	0.95***
	ECG		53.18 (23.03)		29.60 (19.75)		331.81 (163.18)		23.83 (17.50)	
4: 4–8	PPG	54	56.31 (22.45)	0.97***	31.82 (18.19)	0.97***	386.42 (197.17)	0.98***	33.28 (18.27)	0.93***
	ECG		52.52 (22.29)		28.79 (18.08)		393.58 (214.40)		26.47 (15.89)	
5: 5–8	PPG	55	52.99 (18.09)	0.97***	29.79 (16.97)	0.95***	409.77 (213.88)	0.99***	30.37 (19.14)	0.92***
	ECG		48.64 (18.90)		26.25 (17.16)		428.97 (232.79)		24.76 (16.28)	

PPG = photoplethysmography; ECG = electrocardiogram; RMSSD = root mean square of successive RR difference; pNN50 = percentage of successive normal sinus beat-to-beat intervals more than 50 milliseconds; LFms^2 = low-frequency power in milliseconds squared; HFms^2 = high-frequency power in milliseconds squared; M (SD) = mean (standard deviation); ICC = intraclass correlation coefficient.

*** On ln-transformed variables, $p < .001$.

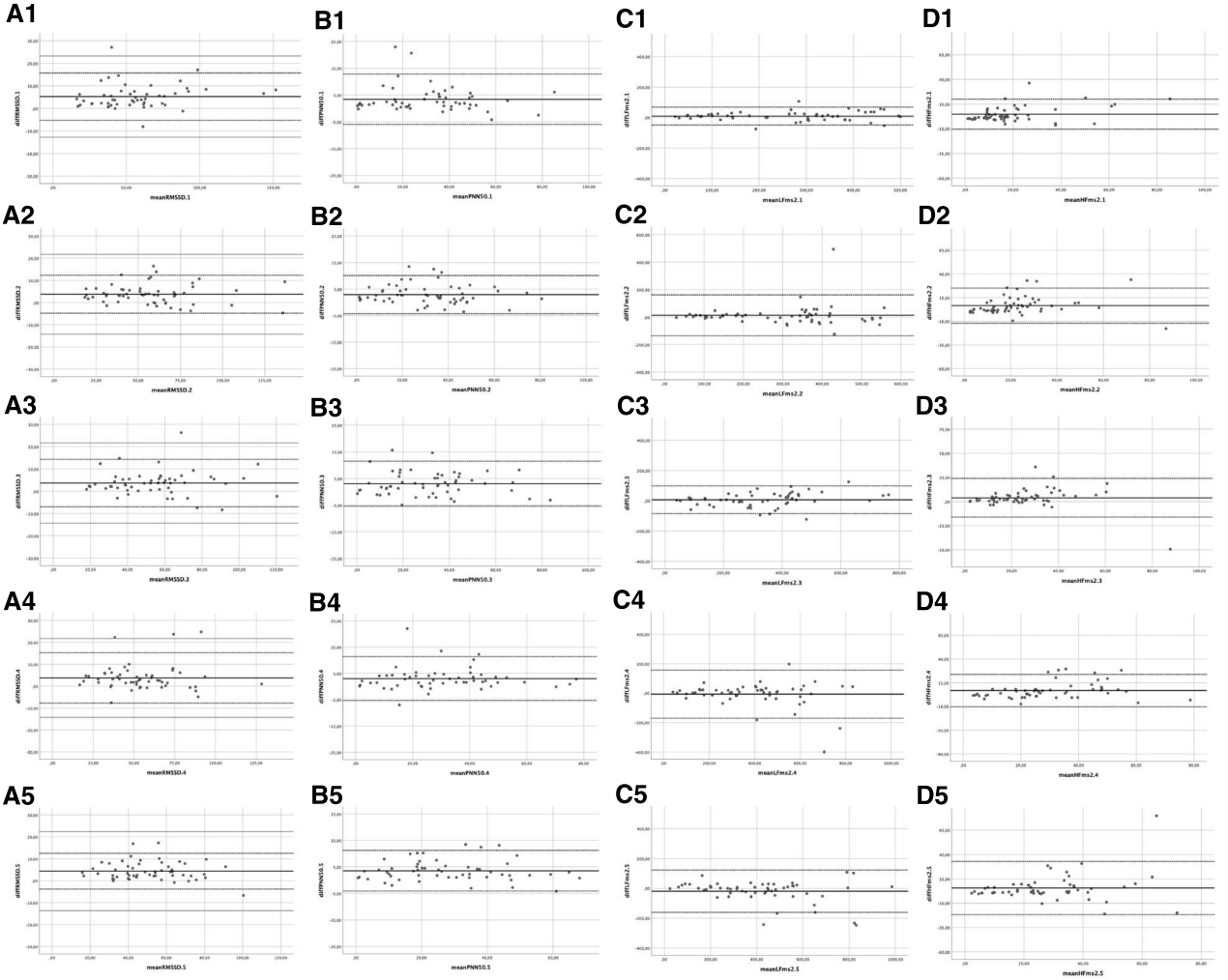

FIGURE 1. Bland-Altman plots comparing PPG with ECG measurements of (A) RMSSD, (B) pNN50, (C) $LFms^2$, and (D) $HFms^2$ per breathing pace (1–5). The solid black lines display the mean bias (constant error), the dotted black lines indicate the upper and lower limits of agreement, and the gray lines in the RMSSD plot indicate the acceptable limits of agreement based on the Task Force (1996). Of note, the acceptable limits of agreement were not plotted for $LFms^2$ and $HFms^2$, as they were far outside the range of the data points. PPG = photoplethysmography; ECG = electrocardiography; RMSSD = square root of mean squared differences of successive normal-to-normal intervals; pNN50 = percentage of successive normal sinus beat-to-beat intervals more than 50 milliseconds; $LFms^2$ = low-frequency power in milliseconds squared; $HFms^2$ = high-frequency power in milliseconds squared.

The quality of the agreement RMSSD, $LFms^2$, and $HFms^2$ was good, as the limits of agreement remained within the a priori defined acceptable values (see Figure 1A for RMSSD). Because the reference in SDs for $LFms^2$ and $HFms^2$ presented in the Task Force article (1996) was substantially larger (SD = 416 and SD = 203, respectively) than the SDs that were found in this study, the acceptable limits of agreement ranged outside the scope of the observations in this study and were therefore not plotted in Figure 1. For all four HRV indices, in accordance with Bland-Altman guidelines, nearly all plots (70%) showed that at least 95% of the observations (i.e., 54–55 observations) lie within 1.96 SD of the mean difference. Five plots (25%) from different indices had 93% of the observations within the range of 1.96 SD, and of only one plot (breathing pace 4 of $HFms^2$), 91% of the observations are within the limits of agreement.

No indications for proportional bias (i.e., a constant difference between the measurements) were found for RMSSD or pNN50, as all trends were small and not statistically significant. There were three exceptions. For $LFms^2$, only for breathing pace 3 (i.e., inhalation 4 seconds and exhalation 7 seconds), statistically significant differences between PPG and ECG for higher $LFms^2$ values were found. For $HFms^2$, statistically significant trends, small to moderate in size, for breathing paces 1 and 4 were found, showing larger differences between PPG and ECG measures for higher $HFms^2$ values. These significant trends suggest proportional bias for these three breathing paces.

Resonance Frequency

Data of 10 participants were missing on one or two breathing paces (total 11 observations); hence, their resonance frequency could not be reliably determined. Therefore, convergence of resonance frequency was based on 47 participants. The distribution of resonance frequency as determined through PPG and ECG is shown in Table 2.

The ICC (absolute agreement, single measure, not allowing for a difference of one between resonance frequencies) showed moderate correspondence in LF peak breathing pace detection between PPG and ECG (ICC = 0.54). PPG and ECG resonance frequency, allowing for a difference of one between breathing paces demonstrating the highest LF peak, corresponded in 36 of 47 participants (76.6%).

Potential Moderators of Resonance Frequency

Of the participants with diaphragmatic breathing experience, 18 of 24 (75.0%) demonstrated corresponding resonance frequency, similar to the percentage found among participants without diaphragmatic breathing experience (18 of 23 [78.0%]).

For the sample of 47 participants, the relative risk ("chance") to obtain resonance frequency correspondence was 0.96 (95% confidence interval [CI] = 0.70–1.31) among participants with diaphragmatic breathing experience compared with those without experience, 0.97 (95% CI = 0.69–1.36) among men compared with women, and 1.01 (95% CI = 0.73–1.39) among participants "low" versus "high" on age, 1.23 (95% CI = 0.88–1.72) among participants "lower" versus "higher" on weight, and 1.20 (95% CI = 0.87–1.66) among participants "lower" versus "higher" on height. χ^2 Tests showed no significant differences in resonance frequency on the variables (see Table S1, Supplemental Digital Content, http://links.lww.com/PSYMED/A954).

Reliability of PPG and ECG

Interbreathing pace reliability (i.e., over all breathing paces) was very large for both PPG- and ECG-derived HRV time-domain measures and large for the frequency-domain measures. ICCs were as follows: lnRMSSD-PPG = 0.90, lnRMSSD-ECG = 0.92, lnpNN50-PPG = 0.89, pNN50-ECG = 0.91, $\ln LFms^2$-PPG = 0.76, $LFms^2$-ECG = 0.78, $\ln HFms^2$-PPG = 0.70, and $\ln HFms^2$-ECG = 0.75.

DISCUSSION

In the current study, we aimed to validate PPG HRV measurements against classical ECG during HRV-BF by comparing HRV indices RMSSD, pNN50, LFpower, HFpower, and resonance frequency between the two methods. Overall, we found that the time-domain and frequency-domain HRV indices can be reliably measured through PPG during the slow-breathing exercises that are used in HRV-BF. Second, for all four HRV indices, large ICCs between PPG and ECG were found, with large effect sizes. Third, Bland-Altman analyses showed good interchangeability of PPG and ECG for RMSSD, LFpower, and HFpower, as indicated by the limits of agreement of all breathing paces falling within the acceptable range for ECG-derived HRV indices, based on the criterion defined by Charlot et al. (42), using the normative data for these HRV indices presented by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). It should be noted that this criterion can be considered rather stringent, and, for example, using normative data from the study of Navarro-Lomas et al. (45), a study with a relatively large and more similar sample with respect to the age range, results in >95% of observations in all breathing paces falling within the acceptable limits of agreement. Fourth, no systematic pattern was found in the breathing paces for which the Task force criterion was not met. Fifth, systematic differences between PPG and ECG measurements of HRV indices were small to very small, further supporting the reliability of the PPG measurements. Finally, no indications were found for proportional bias for any of the HRV indices, although an incidental significant small trend was found for the frequency domain measures in one or two breathing paces. These points show that the Happitech SDK can reliably measure RMSSD, pNN50, LFpower, and HFpower during slow diaphragmatic breathing. With regard to determining resonance frequency, a key procedure in HRV-BF, correspondence between PPG and ECG measurements was good to very good. Individual characteristics, such as experience with diaphragmatic breathing, age, height, weight, or sex, did not affect the PPG-ECG–derived resonance frequency agreement. In sum, our findings provide strong support for PPG as a valid method to measure HRV and to guide HRV-BF breathing exercises.

Despite the strong relative agreement and small differences, we found that PPG-derived time-domain HRV indices were systematically slightly higher than ECG-derived time-domain HRV indices, indicating a small overestimation of PPG measures. An overestimation by PPG is consistent with the literature using various tasks (e.g., cycling) and participant positions, such as sitting or lying down (e.g., (17,42,46)). However, overestimation of time-domain HRV measures by PPG seems more substantial with more sympathetic activation due to physical movement or exercise (e.g., (42)) and less of a threat to the reliability of HRV indices when conducting breathing exercises for HRV-BF interventions. Even though accuracy of beat detection using PPG, which is important for RMSSD and pNN50 assessment, is not perfect, results imply that it is sufficient for HRV assessment during slow breathing.

TABLE 2. Frequency of Highest LF Peak Detection Per Breathing Pace of PPG and ECG Assessment, by Breathing Experience

	PPG			ECG		
Breathing Pace	*n* (%)	*n* (%) Experience	*n* (%) No Experience	*n* (%)	*n* (%) Experience	*n* (%) No Experience
1: 4–5	5 (10.6)	1 (4.2)	4 (17.4)	5 (10.6)	1 (4.2)	4 (17.4)
2: 4–6	6 (12.8)	1 (4.2)	5 (21.7)	9 (19.2)	2 (8.3)	7 (30.4)
3: 4–7	10 (21.3)	7 (29.2)	3 (13.0)	5 (10.6)	5 (20.8)	0 (0.0)
4: 4–8	5 (10.6)	2 (8.3)	3 (13.0)	8 (17.0)	4 (16.7)	4 (17.4)
5: 5–8	21 (44.7)	13 (54.2)	8 (34.8)	20 (42.6)	12 (50.0)	8 (34.8)
Total *n*	47	24	23	47	24	23

LF = low frequency; PPG = photoplethysmography; ECG = electrocardiogram.

The absence of substantial proportional bias is in accordance with previous findings reported by Esco et al. (16), who assessed agreement of PPG- and ECG-derived RMSSD during supine, seated, and standing positions and found no proportional differences between the two methods for any of the positions. Hence, our data further support that being seated is an adequate position to measure PPG-based HRV.

As predicted, we found that resonance frequency had a good level of agreement between PPG-and ECG-derived identification. These findings imply that PPG is a valid method to identify resonance frequency, although there seems to be room for improvement. Excellent, rather than good, agreement would better serve the need for accurate identification of resonance frequency, which is key to performing HRV-BF breathing exercises. It should be noted that, in the HRV-BF treatment protocol, identification of resonance frequency is performed at least twice to improve reliability (12). Repeating the resonance frequency identification procedure allows participants to get acquainted with the procedure, which may be expected to facilitate identification, as diaphragmatic breathing is conducted more properly. Results of the study of Kim et al. (47) showed that the LF peak increases with more training, which may improve discrimination of the breathing paces associated with the highest LFpower for both PPG and ECG. The fact that we did not find better PPG-ECG agreement of resonance frequency identification in individuals with experience in diaphragmatic breathing, due to other activities such as singing lessons, may not be at odds with this idea. The results may either imply that experience in diaphragmatic breathing influences resonance frequency detection through PPG and ECG similarly, or that previous experience is irrelevant in the determination of agreement in resonance frequency. Further research is imperative to draw conclusions on the influence that previous experience in diaphragmatic breathing, or in slow breathing as taught in HRV-biofeedback training, has on the agreement in resonance frequency detection between PPG and ECG.

This study has several noteworthy strengths. It is the first study to investigate the correspondence in resonance frequency detection between PPG and ECG. Moreover, we included a sample of women and men with varying ages and varying on previous experience with diaphragmatic breathing that was large enough to assess individual differences in resonance frequency agreement between PPG and ECG. This varied sample adds to the generalizability of the findings to the general population. A limitation of the current study is that the procedure to assess resonance frequency may have negatively affected the agreement between ECG- and PPG-derived HRV measures. Especially the two slowest breathing paces, which are hard to perform for many people, may have influenced signal quality of PPG and ECG, resulting in an underestimation of PPG-ECG agreement among these participants, lowering the estimates of the total sample. Consequently, PPG-ECG agreement for the two lowest breathing paces may seem to be higher among participants for whom one of these two breathing paces is the resonance frequency.

Future studies need to assess correspondence between PPG- and HRV-derived HRV measures among samples with clinical levels of stress, anxiety, or depression, as differences in cardiovascular function associated with these complaints may affect the convergence between PPG and ECG. In addition, it is useful to further investigate the correspondence between PPG- and HRV-derived HRV measures, once participants' resonance frequency has been identified. Because, during HRV-BF intervention, participants perform their breathing exercises in resonance frequency, determining correspondence between PPG- and ECG-derived HRV indices along the actual intervention would provide further information about PPG's potential to guide HRV-BF breathing exercises. Moreover, it would be valuable to assess whether training improves PPG-ECG correspondence of resonance frequency detection, and if so, which amount of training is required to accurately determine one's resonance frequency. If this information would become available, a training phase could be offered before the actual HRV-BF intervention so that the HRV-BF resonance frequency detection procedure can be improved.

The current study has implications for clinical practice. We suggest that additional research should be conducted on the PPG-ECG agreement of HRV indices in clinical samples before PPG-derived HRV indices are used as an alternative for ECG-derived HRV indices for HRV-BF interventions among individuals with burnout, depression, or anxiety disorders.

In conclusion, using PPG-derived HRV indices, as calculated using the Happitech SDK, is a potentially reliable and valid method for the assessment of HRV during slow, paced breathing. Furthermore, our study is the first to show that PPG can adequately replace ECG assessment to measure resonance frequency during HRV-BF. Future research should investigate how HRV-BF can be offered by using mobile PPG assessments in more naturalistic and clinical settings, which can potentially increase the widespread availability for such a valuable intervention.

Source of Funding and Conflicts of Interest: This work was funded by ZonMW (project number 543003104) awarded to A.C.H. The authors report no conflicts of interest.

REFERENCES

1. Gevirtz R. The promise of heart rate variability biofeedback: evidence-based applications. Biofeedback. 2013;41:110–20.
2. Lehrer P, Kaur K, Sharma A, Shah K, Huseby R, Bhavsar J, et al. Heart rate variability biofeedback improves emotional and physical health and performance: a systematic review and meta analysis. Appl Psychophysiol Biofeedback [Internet] 2020;45:109–29.
3. Penttilä J, Antti H, Jartti T, Kuusela T, Huikuri H, Tulppo M, et al. Time domain, geometrical patterns., and frequency domain analysis of cardiac vagal outflow: effects of various respiratory patterns. Clin Physiol 2008;21:365–76.
4. Appelhans BM, Luecken LJ. Heart rate variability as an index of regulated emotional responding. Rev Gen Psychol 2006;10:229–40.
5. Taelman J, Vandeput S, Spaepen A, Van Huffel S. Influence of mental stress on heart rate and heart rate variability. IFMBE Proc 2008;22:1366–9.
6. Perna G, Riva A, Defillo A, Sangiorgio E, Nobile M, Caldirola D. Heart rate variability: can it serve as a marker of mental health resilience?: Special section on "Translational and neuroscience studies in affective disorders" Section Editor, Maria Nobile MD, PhD. J Affect Disord [Internet] 2020;:754–61.
7. Vanderlei LC, Pastre CM, Hoshi RA, Carvalho TD, Godoy MF. Basic notions of heart rate variability and its clinical applicability. Rev Bras Cir Cardiovasc 2009; 24:205–17.
8. Kemp AH, Quintana DS. The relationship between mental and physical health: insights from the study of heart rate variability. Int J Psychophysiol [Internet] 2013;89:288–96.
9. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018;15:235–45.
10. Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int J Cardiol [Internet] 2010;141:122–31.
11. Vaschillo EG, Vaschillo B, Lehrer PM. Characteristics of resonance in heart rate variability stimulated by biofeedback. Appl Psychophysiol Biofeedback 2006;31: 129–42.
12. Lehrer PM, Vaschillo B, Zucker T, Graves J, Katsamanis M, Aviles M, et al. Protocol for heart rate variability biofeedback training. Biofeedback 2013;41:98–109.

13. Berntson GG, Bigger JT Jr., Eckberg DL, Grossman P, Kaufmann PG, Malik M, et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology 1997;34:623–48.
14. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 2018;9:743.
15. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007;28:R1–R39.
16. Esco MR, Flatt AA, Nakamura FY. Agreement between a smartphone pulse sensor application and electrocardiography for determining lnRMSSD. J Strength Cond Res 2017;31:380–5.
17. Heathers JA. Smartphone-enabled pulse rate variability: an alternative methodology for the collection of heart rate variability in psychophysiological research. Int J Psychophysiol [Internet]. 2013;89:297–304.
18. Vandenberk T, Stans J, Mortelmans C, Van Haelst R, Van Schelvergem G, Pelckmans C, et al. Clinical validation of heart rate apps: mixed-methods evaluation study. JMIR Mhealth Uhealth 2017;5:e129.
19. Claes J, Buys R, Avila A, Finlay D, Kennedy A, Guldenring D, et al. Validity of heart rate measurements by the Garmin forerunner 225 at different walking intensities. J Med Eng Technol 2017;41:480–5.
20. Lu G, Yang F, Taylor JA, Stein JF. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. J Med Eng Technol 2009;33:634–41.
21. Selvaraj N, Jaryal A, Santhosh J, Deepak KK, Anand S. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. J Med Eng Technol 2008;32:479–84.
22. Kuntamalla S, Lekkala RGR. Quantification of error between the heartbeat intervals measured form photoplethysmogram and electrocardiogram by synchronisation. J Med Eng Technol [Internet] 2018;42:389–96.
23. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart-rate-variability recording with smartphone photoplethysmography, polar H7 chest strap, and electrocardiography. Int J Sports Physiol Perform 2017;12:1324–8.
24. van Dijk W, Oosterman M, Jansen I, de Vente W, Huizink A. Stress- and smoke free pregnancy study protocol: a randomized controlled trial of a personalized eHealth intervention including heart rate variability-biofeedback to support pregnant women quit smoking via stress reduction. BMC Public Health 2021;21:1–14.
25. Mol D, Riezebos RK, Marquering HA, Werner ME, Lobban TCA, de Jong JSSG, et al. Performance of an automated photoplethysmography-based artificial intelligence algorithm to detect atrial fibrillation. Cardiovasc Digit Health J 2020;1:107–10.
26. Kochanska G, Kim S. Difficult temperament moderates links between maternal responsiveness and children's compliance and behavior problems in low-income families. J Child Psychol Psychiatry 2013;54:323–32.
27. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Heal 2017;5:258.
28. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39:175–91.
29. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Lawrence Erlbaum Associates, Hillsdale; 1988.
30. Hopkins WG. A scale of magnitudes for effect statistics. A new view Stat 2002; 502:321.
31. Prinsloo GE, Rauch HGL, Karpul D, Derman WE. The effect of a single session of short duration heart rate variability biofeedback on EEG: a pilot study. Appl Psychophysiol Biofeedback 2013;38:45–56.
32. Zaccaro A, Piarulli A, Laurino M, Garbella E, Menicucci D, Neri B, et al. How breath-control can change your life: a systematic review on psycho-physiological correlates of slow breathing. Front Hum Neurosci 2018;12:353.
33. Prinsloo GE, Rauch HGL, Lambert MI, Muench F, Noakes TD, Derman WE. The effect of short duration heart rate variability (HRV) biofeedback on cognitive performance during laboratory induced cognitive stress. Appl Cogn Psychol 2011;25:792–801.
34. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54: 1063–70.
35. Lehrer PM, Gevirtz R. Heart rate variability biofeedback: how and why does it work? Front Psychol 2014;5:5(JUL).
36. Van Diest I, Verstappen K, Aubert AE, Widjaja D, Vansteenwegen D, Vlemincx E. Inhalation/exhalation ratio modulates the effect of slow breathing on heart rate variability and relaxation. Appl Psychophysiol Biofeedback 2014;39:171–80.
37. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol 2014;5: 1040.
38. Lehrer PM, Vaschillo E, Vaschillo B, Lu SE, Eckberg DL, Edelberg R, et al. Heart rate variability biofeedback increases baroreflex gain and peak expiratory flow. Psychosom Med 2003;65:796–805.
39. Widjaja D, Vandeput S, Taelman J, Braeken MAKA, Otte RA, Van Den Bergh BRH, et al. Accurate R peak detection and advanced preprocessing of normal ECG for heart rate variability analysis. Comput Cardiol 2010;2010:533–6.
40. Zheng G, Wang Y, Chen Y. Study of stress rules based on HRV features. J Comput 2018;29:41–51.
41. Bland JM, Altman DG. Statistical methods for assessing agree- ment between two methods of clinical measurement. Lancet 1986;8476:307–10.
42. Charlot K, Cornolo J, Brugniaux JV, Richalet JP, Pichon A. Interchangeability between heart rate and photoplethysmography variabilities during sympathetic stimulations. Physiol Meas 2009;30:1357–69.
43. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–163.
44. Giavarina D. Understanding Bland Altman analysis. Past Present Futur Stat Sci 2014;25:359–72.
45. Navarro-Lomas G, De-La OA, Jurado-Fasoli L, Castillo MJ, Femia P, Amaro-Gahete FJ. Assessment of autonomous nerve system through non-linear heart rate variability outcomes in sedentary healthy adults. PeerJ 2020;8:e10178.
46. Holmes CJ, Fedewa MV, Winchester LJ, Macdonald HV, Wind SA, Esco MR. Validity of smartphone heart rate variability pre-and post-resistance exercise. Sensors (Switzerland) 2020;20:1–13.
47. Kim S, Zemon V, Cavallo MM, Rath JF, Mccraty R, Foley FW. Heart rate variability biofeedback, executive functioning and chronic brain injury. Brain Inj 2013;27:209–22.