

RESEARCH

Open Access



# The effectiveness and user experience of a biofeedback intervention program for stress management supported by virtual reality and mobile technology: a randomized controlled study

Jasmine I. Kerr<sup>1,2\*</sup>, Raphael P. Weibel<sup>1,2†</sup>, Mara Naegelin<sup>1,2</sup>, Andrea Ferrario<sup>1,2</sup>, Victor R. Schinazi<sup>3,7</sup>, Roberto La Marca<sup>4,5</sup>, Christoph Hoelscher<sup>6,7</sup>, Urs M. Nater<sup>8,9</sup> and Florian von Wangenheim<sup>2,7</sup>

## Abstract

**Background** Heart rate variability biofeedback (HRV-BF) can be used for stress management. Recent feasibility studies suggest that delivering HRV-BF in virtual reality (VR) is associated with better user experience (UX) and might yield more beneficial changes in HRV than two-dimensional screens. The effectiveness of a VR-supported HRV-BF intervention program has, however, not been investigated yet.

**Methods** In this study, 87 healthy women and men were assigned to a VR-supported HRV-BF intervention (INT;  $n = 44$ ) or a wait-list control (WLC;  $n = 43$ ) group. The INT came to the lab for four weekly HRV-BF sessions in VR using a head-mounted display. Between lab sessions, participants were asked to perform breathing exercises without biofeedback supported by a mobile application. Stress-related psychological and psychophysiological outcomes were assessed pre- and post-intervention and at a follow-up four weeks after the intervention in both groups. A psychosocial stress test was conducted post-intervention to investigate changes in stress reactivity. UX was assessed after each HRV-BF session in the INT.

**Results** Analysis revealed that LF increased significantly from pre- to post-, whereas pNN50 increased and chronic stress decreased significantly from pre-intervention to follow-up in the INT compared to the WLC. Anxiety and mental fatigue decreased significantly, while mindfulness and health-related quality of life increased significantly from pre- to post- and from pre-intervention to follow-up in the INT compared to the WLC (all small effects). The two groups did not differ in their stress reactivity post-intervention. As for UX in the INT, the degree of feeling autonomous concerning technology adoption significantly decreased over time. Competence, involvement, and immersion, however,

†Jasmine I. Kerr, Raphael P. Weibel and Mara Naegelin contributed equally to this work.

\*Correspondence:

Jasmine I. Kerr  
jkerr@ethz.ch

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

increased significantly from the first to the last HRV-BF session, while hedonic motivation significantly peaked in the second session and then gradually returned to first-session levels.

**Conclusions** This HRV-BF intervention program, supported by VR and mobile technology, was able to significantly improve stress indicators and stress-related symptoms and achieved good to very good UX. Future studies should control for potential placebo effects and emphasize higher degrees of personalization and adaptability to increase autonomy and, thereby, long-term health and well-being. These findings may serve as a first step towards future HRV-BF applications of cutting-edge, increasingly accessible technologies, such as wearables, VR, and smartphones, in the service of mental health and healthcare.

**Trial registration** The study was registered retrospectively as a clinical trial on ISRCTN registry (ISRCTN11331226, 26 May 2023).

**Keywords** Virtual reality, User experience, Biofeedback, Intervention, Heart rate variability, Stress reactivity, Cortisol, Alpha-amylase

## Background

A general shift to digital work over the last decades and recent events such as the COVID-19 pandemic have contributed to the increasing prevalence of stress in our society [1–6]. While acute stress is mostly innocuous, as it reflects an adaptive response to demands and challenges [7, 8], repeated or longer periods of stress (i.e., chronic stress) have been shown to harm physical and mental health [9–11]. Chronic stress is associated with anxiety disorders [12], depression [13], vital exhaustion [14], pain and fatigue syndromes [9], and cardiovascular problems [15], including hypertension [9], among other symptoms and disorders, and negatively affects an individual's psychological well-being and quality of life [16, 17]. The alarming rise of stress and associated symptoms and disorders calls for effective prevention and treatment options. One increasingly recognized stress management method is heart rate variability biofeedback (HRV-BF) [18, 19]. Based on the Association for Applied Psychophysiology and Biofeedback's general definition of biofeedback [20], HRV-BF can be defined as a process of teaching individuals to regulate cardiac activity to improve health-related measures. Immediate and precise feedback on users' cardiac activity, measured using electrocardiography or photoplethysmography devices, combined with changes in thoughts, emotions, and behavior, promotes the desired physiological and psychological improvements. With practice, these changes can become lasting even without the ongoing use of a device [20]. HRV-BF is effective in reducing chronic stress and various stress-related physiological and psychological symptoms such as hypertension, cardiovascular disease, asthma, pain, sleep disturbances, hypertension, cognitive performance, depression, anxiety, and emotional states more generally (for reviews and meta-analyses, see [19, 21–26]). In their meta-analysis on the effectiveness of HRV-BF, Lehrer et al. [22] found medium effect sizes for reducing anxiety and depression and small effect sizes for decreasing perceived stress and improving cardiovascular

measures in studies conducted up until 2018. Pizzoli et al. [23] found even larger effect sizes regarding reducing depressive symptoms – a degree of effectiveness comparable to cognitive behavioral therapy. They note that these differences in meta-analytic effect sizes compared to Lehrer et al. [22] might be due to the inclusion of more recent studies (years 2019–2020) and the use of newer biofeedback (BF) devices able to provide more sophisticated visual feedback making HRV-BF more effective [23].

Heart rate variability (HRV) reflects the variability of time intervals between two heartbeats [27]. HRV at rest indicates the balance of the parasympathetic and sympathetic branches of the autonomic nervous system (ANS) [28]. High resting HRV, indicating high parasympathetic control, has been linked to psychological resilience and the capacity to adapt to changing demands, while low resting HRV, indicating low parasympathetic control, has been linked to chronic stress, psychopathology and increased mortality [28–34]. During HRV-BF, HRV is intentionally maximized through breathing. These increases in HRV are associated with increased vagal afferent transmission to prefrontal regions involved in executive control [35]. When practiced over a longer period, HRV-BF is believed to ameliorate heart-brain connectivity and improve self-regulation, allowing individuals to better cope with stressful situations [30, 36–40]. Indeed, studies have shown that HRV-BF interventions lead to increases in resting HRV [37, 41–43], as well as a buffered decrease in HRV during stress and a greater increase of HRV during recovery compared to control groups [41, 42, 44]. Generally, individuals with higher HRV have also been observed to recover more swiftly from stress regarding cortisol levels [45]. Cortisol indicates hypothalamic-pituitary-adrenal (HPA) axis re/activity and affects several bodily processes important to overcoming stressful situations [46]. The ANS and the HPA axis are anatomically and functionally

connected, whereby the ANS has been found to moderate the reactivity of the HPA axis [45, 47–50]. According to Porges et al. [29], the parasympathetic nervous system regulates an individual's response to stress by modulating the activity of the sympathetic nervous system and HPA axis.

The intentional increases in HRV during HRV-BF are brought about through slow and paced breathing at resonance frequency (RF) [38, 39]. RF can be identified as a high-amplitude peak in the low-frequency (LF; 0.04–0.15 Hz) band of the power spectrum of HRV [51]. Resonant breathing stimulates the baroreflex and maximizes respiratory sinus arrhythmia [51]. The RF of an individual lies between 4.5 and 6.5 breaths per minute and is determined before HRV-BF training [38, 52]. Studies have shown that resonant breathing leads to higher increases in HRV than breathing at an average rate of 6 breaths per minute [53–55]. Initially, clients are often guided by an auditory or visual pacer set to their RF. In time, clients learn to maximize their HRV based on the provided visual feedback of, for example, heart rate (HR) fluctuations or shifts in the power spectrum of HRV. Ultimately, the goal of HRV-BF is to increase parasympathetic and dampen sympathetic activation without the continuous help of supervision, BF, or technological devices [30, 38, 56]. Most HRV-BF intervention programs include daily resonant breathing of at least 10 minutes to reinforce the effects on the baroreflex, parasympathetic nervous system, and emotion-modulating regions of the brain [22–24], and to support the transfer of skills into everyday life [56]. Typically, these HRV-BF intervention programs have been based on simple HR monitors (e.g., cardio tachometer) or other HR tracing devices on computer screens [22, 57].

More recently, the use of virtual reality (VR) technology for BF has been proposed as a novel form of delivery for treating different symptoms and disorders such as pain, post-traumatic stress disorder, or anxiety [58–60], but also for stress management purposes [57, 61]. VR has been investigated because it enables high levels of agency and presence for users in virtual environments due to eliminating external distractions and increased ease of use, especially in combination with head-mounted displays (HMDs) [62]. According to the Cognitive Affective Model of Immersive Learning (CAMIL; [62]), these two main affordances of VR in HMD (i.e., high presence and agency) positively influence factors such as motivation, self-efficacy and self-regulation, which, in turn, foster the effectiveness of learning and training in general [62]. Similarly defined constructs, such as motivation, or related constructs, such as autonomy, mastery and learnability, have also been identified as key properties of an effective biofeedback training by the psychoengineering paradigm

[63]. In line with CAMIL [62], initial empirical evidence indicates that using a VR with HMD for HRV-BF might be especially effective. Studies investigating the feasibility of VR-supported HRV-BF for stress management found short-term improvements in stress-related measures and high levels of motivation, involvement, and self-efficacy [53, 61, 64–70]. Blum et al. [65] found that using an HMD for an HRV-BF session led to significantly higher levels of mindfulness and attention resources and reduced mind-wandering than using a two-dimensional screen. In our work [53], we compared standardized paced breathing to RF HRV-BF delivered both with an HMD and a desktop screen in a single-session study. Results show that using HMDs for HRV-BF and paced breathing reduced psychological stress as much as using desktop screens. Moreover, increases in the coherence ratio (CR; a measure derived from LF) and immersion adaptation were significantly higher while performing HRV-BF and paced breathing in VR than using a desktop screen. Independent of technology, HRV-BF led to higher increases in LF power and was associated with higher degrees of presence than paced breathing. We also found that participants felt competent and autonomously motivated to adopt both technologies and techniques. According to Self-Determination Theory (SDT) [71], satisfying these two basic psychological needs, an individual's competence and autonomy, is important for overall health and well-being, both on the level of technology and in life, in general, [71, 72]. Health interventions that are able to satisfy basic psychological needs have been shown to promote autonomous motivation [73] and improve well-being and reduce stress [74].

However, the mentioned VR-supported HRV-BF interventions for stress management have so far mostly been investigated in single-session studies. There are only two studies investigating the effects of VR-supported HRV-BF interventions over multiple sessions. One study by Gaggioli et al. [64] compared a VR-based intervention program using an HMD that combined stress inoculation training and relaxation with BF to an active control group (i.e., a cognitive behavioral technique intervention) and a wait-list control (WLC) group over 5 weeks. Both interventions were able to reduce perceived stress compared to the WLC, and only the VR-based intervention program was able to significantly reduce anxiety. The study, however, did neither assess any lasting effects in a follow-up assessment nor the user experience (UX) during the intervention itself. Furthermore, due to the combination of both relaxation and stress inoculation, it is not possible to disentangle the specific effects of either method. Lastly, HRV-BF was only used during relaxation in VR to enhance the environment rather than to guide participants'

breathing. In another study, Maarsingh et al. [69] tested the effectiveness and usability of a VR-based stress reduction game using an HMD. Specifically, they assessed the game’s usability in a group of healthy participants in a single session and the effectiveness over three sessions with a clinical sample. They found their application effectively decreased the participant’s stress levels and to be user-friendly with good involvement. The study, however, did not include a follow-up assessment or a control group, and the time between the sessions was not reported.

In summary, initial results look promising and have shown that VR-based HRV-BF interventions delivered with HMDs are feasible. However, existing research is missing the long-term evaluation of effectiveness and the assessment of changes in UX. Investigating specific and lasting psychological and biological effects in longitudinal randomized controlled studies is important to evaluate whether such an intervention would be effective enough to be offered as a program outside of research, for example, for private use at home, at the workplace, or in different healthcare settings. Besides investigating a range of stress-related psychological and psychobiological outcomes, it is also essential to investigate potential changes in UX (e.g., motivation, involvement, and usability) towards the technology over time (e.g., a novelty effect that wears off) as they influence the training outcome [62]. Furthermore, sustained good UX, such as high motivation, is important for adherence to an intervention program [75].

To our knowledge, this is the first study to investigate the effectiveness of using HMD technology, besides the support of mobile technology, to deliver a breathing-based HRV-BF intervention program for stress management. To evaluate this intervention, we aimed to study changes in primary psychological and psychobiological outcomes, secondary psychological outcomes, and UX measures:

- 1) Primary outcomes: Can a VR-supported HRV-BF intervention program improve psychological and psychobiological indicators of stress (re/activity)?
- 2) Secondary outcomes: Can a VR-supported HRV-BF intervention program improve levels of anxiety,

depression, health-related quality of life, fatigue, mindfulness, and psychological well-being?

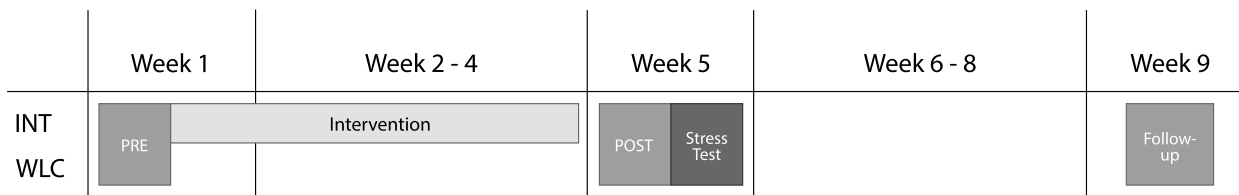
- 3) User experience: What are the UX measures of a VR-supported HRV-BF intervention program, and how do they change with repeated use?

Regarding primary outcomes, we hypothesized that self-reported stress, resting heart rate, and blood pressure would decrease, whereas measures of resting heart rate variability would increase in the intervention compared to the WLC group from pre- to post-intervention and from pre-intervention to follow-up. In addition, we hypothesized that the intervention would lead to a more adaptive stress reactivity in terms of a less heightened and prolonged response compared to a WLC. Regarding secondary outcomes, we hypothesized that self-reported anxiety, depression, fatigue, and frustration of basic psychological needs would decrease while health-related quality of life, mindfulness, psychological well-being, and satisfaction of basic psychological needs would increase in the intervention compared to the WLC group from pre- to post-intervention and from pre-intervention to follow-up. The analysis of UX measures was exploratory, but based on our previous study [53], we expected the user experience to be high at the beginning and to either remain at that level or even improve throughout the study.

**Methods**

**Study design**

In this study, healthy university students were randomly assigned to either participate in an intervention (INT), which consisted of HRV-BF training in VR and using a smartphone application for at-home exercises over four weeks, or to a WLC group (refer to Fig. 1). Stratified by biological sex, the participants were randomly allocated the INT or the WLC. Both groups came to the lab for three measurements. Namely, in week 1 for pre-intervention measurements (*Pre*), in week 5 for post-intervention measurements (*Post*), and in week 9 for follow-up measurements (*Follow-up*). Specifically, participants filled in the same psychological trait questionnaires, had their blood pressure (BP) measured, and had their cardiac activity continuously monitored during all measurements



**Fig. 1** Study protocol. Abbreviations: INT = Intervention group; WLC = Wait-list control group



(i.e., *Pre*, *Post*, *Follow-up*). During *Pre*, all participants answered control and sociodemographic questions and questions on their experience with technology and relaxation methods to assess the characteristics of both groups. After *Post*, both groups underwent a psychosocial stress test immediately after the measurements. Participants in the INT additionally came to the lab once a week from weeks 1 to 4 to undergo the VR-supported HRV-BF intervention. In between lab sessions, participants of the INT performed exercises at home. The first VR-supported HRV-BF session in the lab took place immediately after *Pre*. The WLC received the HRV-BF intervention after *Follow-up*.

### Participants and procedure

Participants were recruited using the University Registration Center for Study Participants [76]. Participants were compensated with 25 Swiss francs per hour of participation (INT: 9.75h; WLC: 3.75h). Potential participants' eligibility regarding the inclusion criteria were assessed during the enrolment using an online screening questionnaire. Participants were required to be fluent in German, between the ages of 18-40, have normal or corrected-to-normal vision, have no disability of arms or hands, and have at least a secondary school diploma. Furthermore, we excluded potential participants that reported acute or chronic somatic diseases, psychiatric disorders, regular use of medication (e.g., antidepressant, antipsychotic, antihypertensive), the consumption of psychoactive substances in the last three months, or heavy drinking ( $\geq 15$  and  $\geq 8$  drinks per week for men and women, respectively [77]). Further, the exclusion criteria included regular tobacco consumption ( $> 5$  cigarettes per week) unless the consumption was limited to the weekends. Women with an irregular menstrual cycle (i.e., not between 27-32 days  $\pm 4$  days), using hormonal contraceptives, that were pregnant or lactating were also excluded from the study due to the respective effects on the ANS and, in particular on HRV (see, e.g., [78–80]).

Due to a lack of consensus and well-established power analysis methods for the statistical models used in the analysis (i.e., multilevel models), we determined the sample size based on propositions concerning the number of participants for randomized controlled trial pilot studies [81] and studies analyzed with multilevel models, which recommend a minimum total sample size of 70 participants [82–84]. Furthermore, we increased the number of participants to account for possible no-shows, dropouts, and missing data. Thus, we aimed to enroll a total of 100 participants (50% female).

To accommodate the planned number of participants, we organized multiple session slots in week 1 of, at most, 15 participants per slot. Participants were free to select a

series of slots that best fitted their schedule for the duration of the study on a first-come-first-served basis; that is, they were asked to return to the lab on the same weekday and time for each week of measurements and/or HRV-BF training. Before participants chose their slot, each slot was randomly assigned a condition (i.e., INT or WLC). To ensure the randomization of participants, this information was concealed and not disclosed to participants before they chose their slots. After choosing a series of slots, participants were informed about their assigned condition. At this point, participants in the WLC were informed that they would only participate in the measurement sessions, and their schedule was adapted accordingly (i.e., pre-scheduled VR-supported HRV-BF sessions were canceled).

### Intervention

The four-week intervention consisted of four weekly VR-supported HRV-BF training sessions at the lab, which were accompanied by psychoeducational videos and breathing exercises outside of the lab, which were guided by a smartphone application. In the first HRV-BF session (S1), participants spent 26 minutes in VR, whereas in sessions 2 to 4 (S2, S3, S4), they spent 20 minutes in VR (refer to Table 1 for detailed information on the session timings). For their homework, participants were asked to perform the breathing exercises for at least 10 minutes on days when they were not in the lab (i.e., six days per week).

### Virtual reality heart rate variability biofeedback

Lehrer et al.'s manuals informed the number and content of intervention sessions [38, 85]. The more recent manual proposes a training period in the laboratory of around 4-5 weekly sessions, including psychoeducational, accompanied by at-home exercises between sessions. Sessions usually last around 25 minutes with 2-minute

**Table 1** Timings for each intervention session at the lab

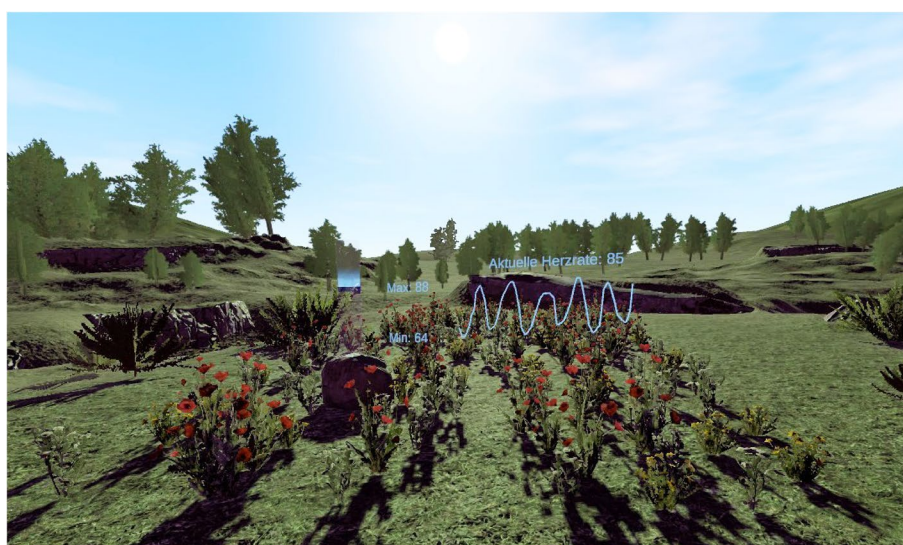
Session 1		Sessions 2-4	
Task	Time	Task	Time
Psychoeducation	20'	Psychoeducation	5'
HMD Handling	5'	HRV-BF block 1	10'
RF determination	6'	Break	2'
Break	2'	HRV-BF block 2	10'
HRV-BF block 1	10'	UX questionnaires	5'
Break	2'		
HRV-BF block 2	10'		
UX questionnaires	15'		

Abbreviations: HMD Head-mounted display, RF Resonance frequency, HRV-BF Heart rate variability biofeedback, UX User experience, 'minutes

breaks in-between biofeedback blocks. In our study, each HRV-BF session also started with a psychoeducational video participants watched on a computer screen. The videos explained the causes and effects of stress on mental and physical health and how BF may help to improve autonomic functioning, gas efficiency of the lungs, and stress management skills. In addition, the videos introduced the individual components of HRV-BF and the meaning of the different visual elements shown in VR and gave instructions on how to breathe. Refer to Table A4 in the supplementary information for a more detailed overview of the content of the videos. Following each psychoeducational video, participants then put on the HMDs to start the HRV-BF training. In S1, participants first completed a six-minute block of paced breathing at 0.25 Hz, from which their individual RF was determined following Sakakibara et al.'s protocol [52]. This protocol offers a more precise and time-efficient estimation of an individual's RF than the protocol proposed by Lehrer et al.'s manuals [38, 86]. Sakakibara's protocol uses the peak frequency in the LF region of HRV under respiratory control at 0.25 Hz as an estimate for individual RF. RF is calculated as the argmax in the region of 0.075 Hz to 0.10833 Hz (i.e., between 4.5 and 6.5 breaths per minute) of the power spectral density from the last 5 min of collected inter-beat intervals (i.e., R-R intervals; RRi). Following RF determination, Lehrer et al. [86] suggest instructing clients to follow a pacer set to their RF in the first two sessions. From the third session onward, clients are instructed to follow the fluctuations in their heart rate to guide RF breathing and only use the pacer when

needed. During the HRV-BF training, participants in our study were asked to breathe slowly and regularly at the pace of their determined RF. In S1, they were guided by an auditory and visual pacer, which was placed as a three-dimensional object on a rock in the field in front of the participants in the VE (refer to Fig. 2). After acclimatizing to the VE and the pacer, real-time HR feedback was introduced during the second block of S1. Specifically, the HR values of the last 25 seconds were shown using a line graph, which was in a fixed position in the VE (i.e., to the right of the pacer; refer to Fig. 2). In S2 and S3, the audiovisual pacer was only present for two minutes at the beginning of each block. In S4, the pacer was omitted completely. In the sessions where the pacer was no longer present, participants were instructed to exhale when the HR reached its highest value and inhale at its lowest value, thus amplifying the sinusoidal HR curve caused by RF breathing. The aim of gradually removing the pacer was to increase participants' self-awareness, self-regulation, feelings of autonomy, and competence and to shift from external to increasingly internal feedback.

The virtual environment (VE) used in both the HRV-BF training and the RF determination was an improved version of our previously developed and tested VE [53]. The VE depicted an open alpine region on a summer day and was experienced from the perspective of sitting on a large tree trunk and overlooking a flat, green field (refer to Fig. 2). Bird songs and the sounds of water flowing in a tiny creek made up the soundscape. With respect to the version used in our previous study [53], we improved the environment by adding new elements wished for by a



**Fig. 2** The virtual environment from the user's perspective. The pacer was placed above a stone on the left side of the meadow. The blue graph in the center was used for immediate feedback, visualizing the heart rate fluctuations of the last 25 seconds. The flowers in the meadow and the sun in the background were used for the heart rate variability cumulative feedback

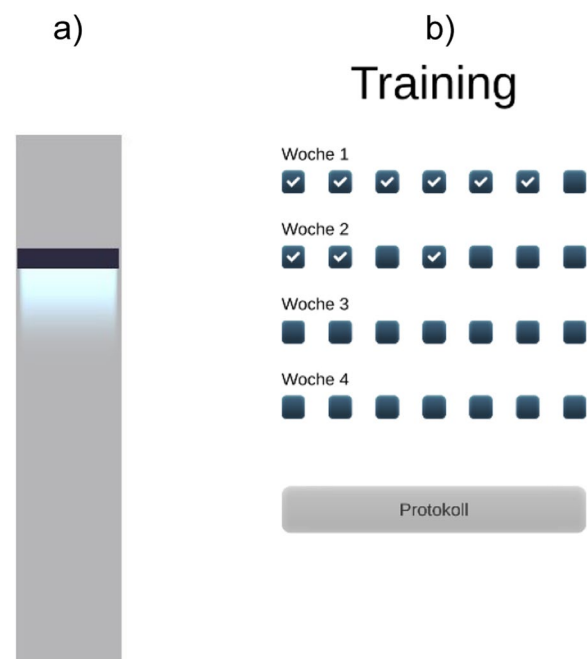
majority of participants (i.e., dynamic clouds and changing sky color over time), polishing the soundscape (i.e., removing sounds that were perceived as annoying by participants and adjusting the volume), moving elements such as the pacer to increase their visibility and optimizing visual performance.

In addition to the immediate feedback, participants received cumulative feedback during the HRV-BF blocks. The cumulative feedback was used to provide participants with easy-to-understand feedback on their overall progress and for positive reinforcement [63, 68]. Based on participants' HRV, their coherence ratio (CR) was computed every 10 seconds from a time frame of the last 90 seconds. The CR is computed based on the relative power of the main peak in the low-frequency band of the HRV signals (i.e., CR is defined as peak power / (total spectral power - peak power) [36]. If the computed CR was  $> 1$  the cumulative feedback score was increased by one point. Otherwise, it remained the same. The VE changed in response to increases in this score. Blossoms grew in the green meadow, the sun rose further on the horizon, and the natural soundscape became more intense. The cumulative score and associated environmental changes were present during all sessions. At the end of each session block, participants were informed about their achieved cumulative score reflecting their progress.

The virtual environment was created using Unity (Unity Technologies, San Francisco, CA, USA) and delivered using an HMD (i.e., Oculus Quest 1, Reality Labs, Menlo Park, CA, USA). The virtual soundscape was heard directly through the HMD's built-in speakers. A wearable electrocardiogram device, the Polar H10 chest belt (Polar Electro Oy, Kempele, Finland), was used to record inter-beat intervals (i.e., R-R intervals; RRi). The recorded RRi values were streamed from the Polar devices to a desktop computer in the lab, using a custom-made Windows application (based on the Polar software development kit [87]). The RRi data were subsequently sent from the desktop computer to a centralized database, where all of the participants' information was stored. The RF and cumulative feedback score were then calculated from these RRi values using a self-developed Python script. The results were stored in the centralized database, from which the VR application regularly polled the latest score and RRi values to effectuate changes in the VE and to compute and visualize the HR graph.

#### Smartphone homework

A self-developed smartphone application supported participants' daily practice outside of the lab setting (refer to Fig. 3). Daily exercises outside of the laboratory are an integral part of the majority of HRV-BF interventions



**Fig. 3** **a** The pacer shown in the mobile application was used for breathing exercises outside of the lab. **b** The progress tracker page of the mobile application allowed participants to keep track of their breathing exercises outside of the lab

[88] and are meant to increase the intervention effects and support skill transfer into daily life [38]. In the app, participants could start a breathing exercise session featuring the same audiovisual breathing pacer from the VE, set to their individual RF. Before each breathing exercise session, participants could choose to practice for either 10 or 20 minutes. Lehrer et al.'s manual [38] recommends asking participants to practice for 20 minutes twice a day. To counteract dropout due to potentially high perceived workload and to increase motivation, we asked participants to at least practice for 10 minutes daily but ideally for longer and more than once a day. Adherence was further fostered by instructing participants to set a daily reminder on the smartphone, asking them how the at-home practice was going after each VR session at the lab, and by an in-app progress tracker (refer to Fig. 3). The app recorded the start and end times of each exercise locally, allowing the participants to keep track of their exercises. White tick marks appeared for each day where participants had practiced for at least 10 minutes. Additionally, the data was sent to a database for later analysis. The application was developed for both iOS and Android using Unity.

#### Stress test

After regular measurements during the *Post* assessment in week 5, participants of both groups additionally



underwent the Trier Social Stress Test for Groups (TSST-G; [89]) to assess participants' stress reactivity. Specifically, participants were subjected to a mock job interview followed by a mental arithmetic task. Both tasks were conducted in the presence of two judges in lab coats and two mock cameras, as well as the other participants in that slot. The judges were instructed to behave in a reserved and neutral manner. The tasks were preceded by a preparation phase of five minutes. In this anticipation phase, participants were informed of the task using a pre-recorded audio message and asked to prepare themselves for the mock job interview. During the mock job interview (12 minutes), each participant was asked to present themselves to the judges and speak freely about why they would be a good candidate (up to two minutes each). In the following arithmetic task (eight minutes), each participant was asked to count backward from 2043 in steps of 17 in front of the judges and the other participants (60 seconds per participant; the next participant continued from the number where the previous participants had left off to avoid learning effects, no group of participants arrived at numbers < 1000). A research assistant was present during the stress test, assisting the panel of judges with the timing and the participants with the second saliva sampling. The stress test was followed by a recovery period, where participants watched a nature documentary for 30 minutes. The stress tests all took place in the afternoons, starting either at 13:30 or at 16:30.

### Outcomes

The primary outcomes of the study were psychological and psychobiological measures of stress: self-reported levels of stress (i.e., periods of one month and one week), HR, HRV, and systolic (SYS) and diastolic (DIA) blood pressure at rest. Additionally, the following stress reactivity measures were used as primary outcomes: self-reported psychological state, levels of salivary cortisol (sCort) and alpha-amylase (sAA), HR, and HRV. Secondary outcomes included levels of anxiety, depression, mindfulness, psychological well-being, health-related quality of life (HRQoL), and fatigue.

### Psychological assessment

For the psychological assessment, we used the following questionnaires during *Pre*, *Post*, and *Follow-up*. The Perceived Stress Scale (PSS-10; for which the required license was obtained; [90, 91]) to assess levels of chronic stress during the last month, with scores ranging from 0 to 40. Stress levels during the last week were assessed using the stress scale of the Depression Anxiety and Stress Scales (DASS-21; [92, 93]), while the remaining scales were used to determine levels of anxiety and depression, with scores of all scales ranging from 0 to 42.

The Mindfulness Attention and Awareness Scale (MAAS; [94]) was used to measure levels of mindfulness, with scores ranging from 1 to 6. The World Health Organisation-Five Well-Being Index (WHO-5; [95, 96]) was used to measure psychological well-being, with scores ranging from 0 to 100, and the Short Form Health Survey (SF-36; [97]) to determine HRQoL. The SF-36 consists of eight scales (i.e., general health, mental health, bodily pain, vitality, social functioning, physical functioning, role-physical, role-emotional, and vitality), with scores ranging from 0 to 100. Higher scores indicate higher HRQoL. The scales role-emotional and role-physical describe role limitations while performing daily activities and work due to difficulties with mental or physical health. In addition, the Multidimensional Fatigue Inventory (MFI-20; [98]) was used as a measure of fatigue, including scales on general fatigue, mental fatigue, reduced activity, and reduced motivation, with scores ranging from 4 to 20. Lastly, the Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS; [99]), based on SDT [71], was used to study whether the training led to satisfaction and/or frustration of basic psychological needs (i.e., autonomy, competence, and relatedness), resulting in six scales with scores ranging from 4 to 20. In addition, the BPNSFS comprises two composite scores reflecting general satisfaction and frustration with these needs, with scores ranging from 12 to 60.

### Physiological assessment

During *Pre*, *Post*, and *Follow-up*, participants' cardiac activity was measured. From the recorded R-R-interval (RRi), we derived HR and HRV following established measurement standards [100]. For the analysis, we used 10 minutes of the data recorded while participants filled in questionnaires (starting one minute after they began the questionnaires to allow for acclimatization). The raw RRi were filtered for motion artifacts and ectopic beats using the threshold-based filtering method from the Python package `hrv`, resulting in the exclusion of three measurements (from three different participants) with >10% noisy data. From the filtered data we extracted the mean HR, the Root Mean Square of Successive Differences (rMSSD), the percentage of successive normal-to-normal intervals differing by >50 ms (pN50), and the standard deviation of normal-to-normal heartbeat intervals (SDNN). Moreover, we extracted the power in the low and high bands (LF and HF) of the power spectral density, and the ratio of LF to HF (LF/HF ratio). The power spectral density was estimated via Fast Fourier Transform from the RRi series interpolated at 4Hz with cubic splines. Baseline measurements of SYS and DIA were assessed using a BP monitor (HBP-1120, Omron, Kyoto, Japan) whilst participants filled in questionnaires



during *Pre*, *Post*, and *Follow-up* (i.e., three measurements per participant). Due to a technical issue, the cardiac and BP measurements of three participants were not recorded at *Follow-up*.

### Stress reactivity

*Psychological state measures* were repeatedly assessed before, during, and after the stress test (-5, 0, +25, and +55 minutes) with the Multidimensional Mood State Questionnaire (MDMQ; which was obtained from the questionnaire's publisher; [101]) and a self-developed Visual Analog Scale (VAS) asking how stressed the participants felt (slider from 0-100). The MDMQ includes the scales MDMQ<sub>mood</sub> ("good mood–bad mood"), MDMQ<sub>calmness</sub> ("calmness–nervousness"), and MDMQ<sub>wakefulness</sub> ("wakefulness–sleepiness"), with scores ranging from 4 to 20. Lower scores on the MDMQ scales indicate bad mood, nervousness, or sleepiness.

*Cardiac measures* were continuously recorded during both *Post* and the subsequent stress test, allowing us to analyze the cardiac measures at different intervals before, during, and after the stress test. Specifically, data was extracted in five-minute intervals at baseline before the stress test (TSST<sub>BL</sub>, starting six minutes after participants' began answering the *Post* measurement questionnaires), during the mock job interview (TSST<sub>I</sub>, i.e., +1 to +6 minutes from the start of the stress test), during the arithmetic task (TSST<sub>M</sub>, +13 to +18 minutes), and twice during the recovery phase (TSST<sub>REC1</sub> and TSST<sub>REC2</sub>, i.e., +24 to +29 and +39 to +44 minutes). The filtering of the raw cardiac data and the extracted measures were the same as described in the Physiological assessment section above.

*Biochemical measures* were collected through saliva samples using SaliCaps® (IBL-Tecan, Hamburg, Germany) at four time points (-5, +12, +25, +55 minutes) to analyze sCort (as an indicator of the HPA axis) and sAA levels (as an indicator of the ANS) before during and after the stress test. Fresh samples were stored at -20° Celsius until shipment to the biochemical laboratory of the Faculty of Psychology at the University of Vienna (Vienna, Austria). There, saliva samples were stored further at -30° Celsius until analyses. Salivary cortisol concentration was assessed using a commercial luminescence immunosorbent assay (IBL-Tecan, Hamburg, Germany). Salivary alpha-amylase activity was determined using a kinetic colorimetric test and reagents obtained from DiaSys Diagnostic Systems (Holzheim, Germany) after saliva samples were diluted at 1:400 using 0.9% saline solution [102]. Intra- and inter-assay coefficients of variation were < 10% for both sCort and sAA. For one sample, the concentration of sAA was under the detection threshold of <3 U/mL.

### User experience

Participants in the INT were asked to assess the UX of the intervention each time they used the VR application in the lab (S1-S4). Specifically, they completed the Presence Questionnaire (PQ; [103]), the Flow Short Scale (FSS; [104]), the System Usability Scale (SUS; [105]), four sub-scales of the Unified Theory of Acceptance and Use of Technology questionnaire (UTAUT; [106]), and the Simulator Sickness Questionnaire (SSQ; [107]) at the end of each HRV-BF session. The PQ assesses participants' involvement (score range: 1 to 77), the sensor fidelity (1 to 35), participants' ability to immerse and adapt (1 to 56), and the degree of negative effects of the interface quality (1 to 21). The FSS assesses participants' flow experience (10 to 70) and the degree of anxiety (i.e., fear of failure while using a system; score range 1 to 21). The UTAUT assesses effort expectancy, facilitating conditions, behavioral intention, and hedonic motivation (all scores ranging from 1 to 7). The SSQ assesses the degree of nausea (1 to 467), its effects on the oculomotor systems (1 to 371), and the amount of disorientation felt by participants (1 to 682). Participants were also allowed to give free text answers to address any issues or suggestions not covered by the questionnaires. During *Post* (i.e., after the intervention), participants were asked a set of self-developed questions about using the mobile application for the at-home exercises. In addition, participants were asked to fill in the Autonomy and Competence in Technology Adoption questionnaire (ACTA; [72]) during *Pre* and *Post*, also a measure rooted in SDT [71]. The ACTA addressed how much a person thinks using a certain technology will increase their sense of autonomy and competence using four indices. The Autonomy Regulation Score (ACTA<sub>ARS</sub>) captures intrinsic ("It is going to be fun to use") and identified regulation ("I believe it could improve my life"), and the Controlled Regulation Score (ACTA<sub>CRS</sub>) captures introjected ("I want others to know I use it") and external regulation ("I feel pressured to use it"). Both indices have a score range from 6 to 30. The Relative Autonomy Index (ACTA<sub>RAI</sub>) subtracts ACTA<sub>CRS</sub> from ACTA<sub>ARS</sub>. Finally, the Perceived Competence Score (ACTA<sub>PCS</sub>) measures perceived competence towards a technology, with a score ranging from 3 to 15. Higher scores on all scales reflect higher values of the measured construct.

### Statistical analysis

Data analysis was performed using R (version 4.2.1) and RStudio. A significance level of .05 was used for all tests, and effect size Pearson's *r* was used to quantify the effect sizes of planned contrasts (small: 0.1 to 0.3; medium: 0.3 to 0.5; large: 0.5 or greater; [108]). To investigate the general effectiveness of the intervention concerning the

psychological and psychophysiological outcomes, we ran linear multilevel models with the R package `nlme` with the between-subjects factor *Group* (two levels: INT vs. WLC) and the within-subject factor *Time* (three levels: *Pre*, *Post* and *Follow-up*). Specifically, we were interested in the interaction effects of *Time* x *Group* on psychological and physiological outcomes. To capture any significant changes in the INT from *Pre* to *Post* and from *Pre* to *Follow-up* compared to the WLC, we specified planned contrasts for the factor *Time* using dummy-coding (*Pre* [0,0], *Post* [1,0], and *Follow-up* [0,1]). Using planned contrasts counters the inflation of the Type 1 error [109]. Psychological and psychobiological stress reactivity assessed after the *Post* measurements was also analyzed using a multilevel model with the between-subjects factor *Group* (two levels: INT vs. WLC) and the within-subject factor *Time* (four levels for psychological state and biochemical measures and five levels for cardiac measures). UX measures assessed at the end of S1-S4 (PQ, FSS, SUS, UTAUT, SSQ) were analyzed using multilevel models with the within-subject factor *Time* (four levels: S1, S2, S3, and S4). Post hoc tests were run to capture the effects of time (i.e., comparing S1 to S4 and S2 to S3) and changes to the feedback (i.e., comparing S1 to S2, S2 to S4, and S3 to S4). Pairwise comparisons for these measures were computed using the R package `emmeans` and corrected using the Benjamini-Hochberg method. For the ACTA, which was only assessed at *Pre* and *Post*, we ran two-tailed dependent samples t-tests to explore whether scores had significantly changed from *Pre* to *Post* in the INT. All specified multilevel models were checked for level-one heteroscedasticity, level-one and level-two residuals normality, and multicollinearity (where applicable). We found that the normality assumption for level-one and level-two residuals was violated (i.e., longer tails) by some models. However, it has been demonstrated that fixed effects estimates, which were used for our hypothesis tests, are robust to violations of the normality assumption [110]. Similarly, we found that level-two residuals for random intercepts suggested slight deviations from normality for some models. Regardless, it has been shown that estimates of fixed effects and their standard errors are robust to non-normal level-two residual errors when sample sizes are as large as the one used in this study [111]. Additionally, we checked for influential data points using the `influence.ME` package in R. Here, results revealed that no participant had an undue influence on the model (all Cook's Distances < 1).

## Results

### Study sample

In total, two hundred and three ( $n = 203$ ) participants were assessed for eligibility. One hundred and four

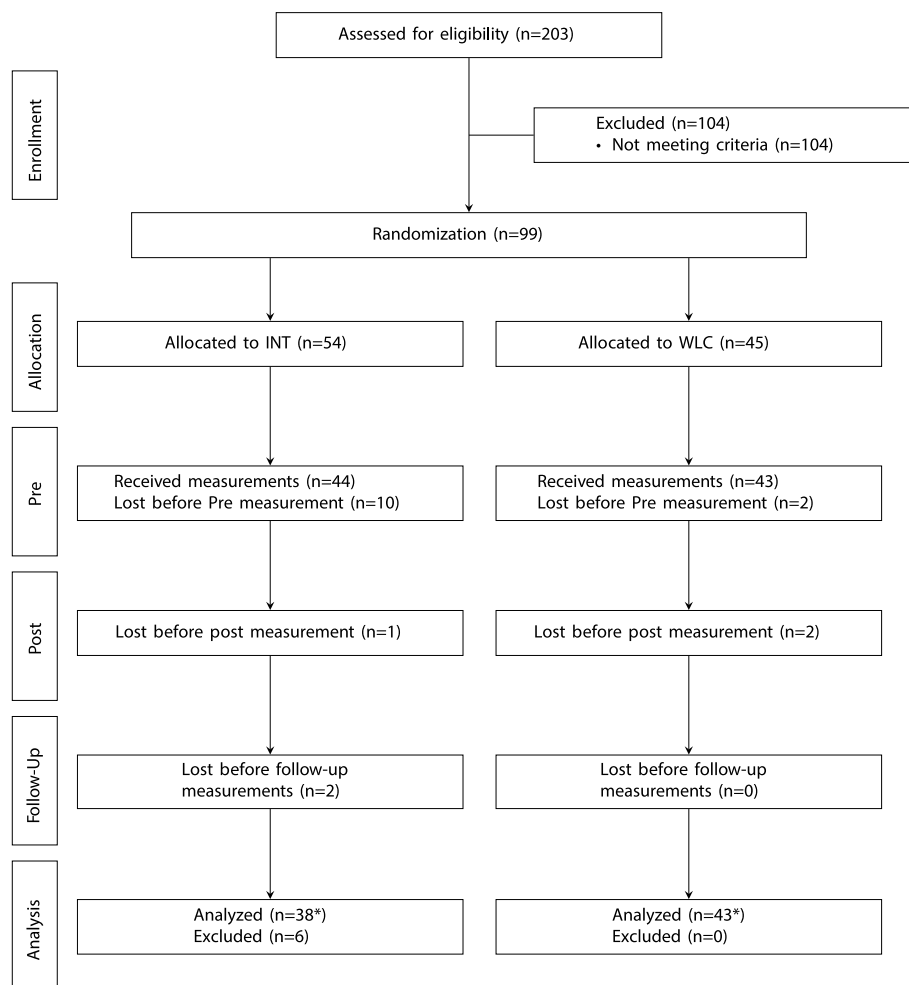
( $n = 104$ ) were excluded, resulting in 99 participants who were randomly assigned to either the INT ( $n = 54$ ) or the WLC ( $n = 45$ ) by blinded slot series sign-up. Of participants assigned to the INT, 44 attended the first session and received the intervention. In the WLC, 43 participants showed up to the first session. Thus, a total of twelve participants were lost before *Pre* measurement. Across both groups, three participants dropped out before *Post* measurements and two dropped out before the *Follow-up* measurement. Incomplete data of seven participants were included in the final analysis. Participants in the INT who took part in all the lab sessions and practiced the breathing exercises at home at least 50% of the time were included in the analysis. Therefore, six participants of the INT were excluded from all analyses, except the analysis of UX measures, because they either did not practice the exercises often enough ( $n = 3$ ) or called in sick ( $n = 3$ ). Finally, the data from 81 participants were analyzed. Refer to Fig. 4 for a participant flow diagram.

The mean age of participants was 22.88 ( $SD = 4.02$ ) (refer to Table 2 for detailed sample characteristics per group). Most participants were university students and had at least a Baccalaureate degree. The majority reported being physically active  $\leq 5$  hours a week, did not smoke, and had no children. A little less than half of the participants reported being in a relationship at the time (33/87). Furthermore, most had either never used an HMD (47/87) or used it less than once a month (37/87). Similarly, participants had no experience with BF except for two people in the INT. Experience with breathing exercises or other exercises for stress management and relaxation was more common in both groups. Analysis after four weeks of intervention showed that participants in the INT practiced 7.7 min/day ( $SD = 1.9$ , range = 0 - 20 min) at home on average across all weeks. The log data also revealed that there was a decrease in the average number of exercises recorded per week from 5.5 ( $SD = 1.32$ ) in the initial week to 4.5 ( $SD = 1.04$ ) in the final week. Of a total of 888 possible in-app exercises, the 37 participants of the INT initiated 732 exercises (82.4%) throughout the four-week intervention. Of these 732 exercises, 92.1% (674) were completed ( $\geq 10$ -minute practice), and 7.9% (58) were aborted ( $< 10$ -minute practice).

### Intervention effects

#### Psychological outcomes

Following up on significant interaction effects, planned contrasts revealed a small yet significant decrease of chronic stress (i.e., PSS-10) from *Pre* to *Follow-up*, and of anxiety and mental fatigue from *Pre* to *Post* and from *Pre* to *Follow-up* in the INT compared to the WLC.



**Fig. 4** Diagram of participant flow. Abbreviations: INT = Intervention group; WLC = Wait-list control group. \* Incomplete data of seven participants were included in the analysis

Moreover, there was a small but significant increase in mindfulness, role-emotional, and social functioning from *Pre* to *Post* and from *Pre* to *Follow-up* in the INT compared to the WLC. No other significant effects were found (refer to Table 3 for detailed results and Tables A1-A2 in the supplementary information for means and standard deviations).

#### Physiological outcomes

Planned contrasts following significant interaction effects for cardiac measures showed a small yet significant increase in LF from *Pre* to *Post* in the INT compared to the WLC and a small yet significant increase in pNN50 from *Pre* to *Follow-up* in the INT compared to the WLC. No other significant effects were found (refer to Table 3 for detailed results and Table A3 in the supplementary information for means and standard deviations).

#### Stress reactivity

With regard to psychological state measures, results of the stress test (after *Post*) showed significant main effects of *Time* on MDMQ<sub>mood</sub>,  $\chi^2(3) = 102.42$ ,  $p < .001$ , MDMQ<sub>calmness</sub>,  $\chi^2(3) = 159.64$ ,  $p < .001$ , and VAS<sub>stressed</sub>,  $\chi^2(3) = 143.50$ ,  $p < .001$ , but not on MDMQ<sub>wakefulness</sub>,  $\chi^2(3) = 2.23$ ,  $p = .53$ . Refer to Fig. 5 for plots of mean changes of MDMQ<sub>mood</sub> and MDMQ<sub>calmness</sub>. As for cardiac measures, there was also a significant effect of *Time* on HR,  $\chi^2(4) = 346.91$ ,  $p < .001$ , rMSSD,  $\chi^2(4) = 120.43$ ,  $p < .001$ , SDNN,  $\chi^2(4) = 129.84$ ,  $p < .001$ , pNN50,  $\chi^2(4) = 118.07$ ,  $p < .001$ , HE,  $\chi^2(4) = 61.55$ ,  $p < .001$ , and LF,  $\chi^2(4) = 60.52$ ,  $p < .001$ , but not on LF/HE,  $\chi^2(4) = 3.42$ ,  $p = .49$ . Finally, analysis of biochemical measures also revealed a significant main effect of *Time* on both sCort,  $\chi^2(3) = 113.74$ ,  $p < .001$ , and sAA,  $\chi^2(3) = 109.58$ ,  $p < .001$ . No other significant interaction effects *Time* ×

**Table 2** Sample characteristics of the intervention group (INT) and wait-list control group (WLC)

Variable	INT (n = 44)	WLC (n = 43)
Sex		
Female	23	19
Male	21	24
Age [M (SD)]	23.82 (4.69)	22.47 (3.45)
Education		
Secondary school degree	1	1
Baccalaureate	28	26
Bachelor degree	10	11
Master degree	5	5
Physical activity		
≤5 hours a week	40	29
>5 hours a week	4	14
Smoking (>5 cigarettes a day)		
Yes	1	2
No	43	41
Relationship status		
Single	26	24
In a relationship	16	17
No answer	2	2
Children		
Yes	2	0
No	42	43
Experience with the use of head-mounted displays		
Never	21	26
Less than once a month	22	15
Monthly	1	1
Weekly	0	1
Experience with relaxation and stress management methods		
None	8	11
Tried briefly	21	16
Practised now and then	13	13
Practised regularly	2	3
Experience with biofeedback		
Yes	2	0
No	42	43
Breathing exercise practice		
Never	12	16
Less than once a month	19	14
Monthly	4	5
Weekly	9	5
Daily	0	3

Group were found (all  $p > .05$ ). Refer to Fig. 5 for plots of mean changes of HR, LF, sAA, and sCort. Taken together, the TSST induced a temporary worsening of mood while increasing subjective stress, HR, sCort, and sAA during the stress test.

**User experience**

User experience, which was only assessed for the INT, revealed a significant and a moderate decrease in the scores of ACTA<sub>ARS</sub> from *Pre* ( $M = 23.51, SD = 2.94$ ) to *Post* ( $M = 21.87, SD = 3.71$ ),  $t(36) = 2.57, p = .01, r = .39$ , and a significant and moderate increase in ACTA<sub>CRS</sub> from *Pre* ( $M = 9.19, SD = 2.41$ ) to *Post* ( $M = 10.60, SD = 3.76$ ),  $t(36) = -2.13, p = .04, r = .33$ . As a result, there was also a significant and large decrease in the scores of ACTA<sub>RAI</sub> from *Pre* ( $M = 14.32, SD = 3.75$ ) to *Post* ( $M = 11.27, SD = 4.17$ ),  $t(36) = 4.41, p < .001, r = .59$ . Finally, the analysis revealed a significant and moderate increase in ACTA<sub>PCS</sub> from *Pre* ( $M = 10.73, SD = 2.10$ ) to *Post* ( $M = 11.68, SD = 1.55$ ),  $t(36) = -2.86, p = .01, r = .43$ .

Concerning changes across all four intervention sessions (S1-S4) of the INT (refer to Table A4 in the supplementary information for detailed results), the analysis revealed a significant main effect of *Time* on PQ<sub>Invo</sub>,  $\chi^2(3) = 25.63, p < .001$ , PQ<sub>SensFi</sub>,  $\chi^2(3) = 10.87, p = .01$ , PQ<sub>ImrsAdpt</sub>,  $\chi^2(3) = 10.46, p = .02$ , UTAUT<sub>HedMotv</sub>,  $\chi^2(3) = 13.57, p = .004$ , SUS,  $\chi^2(3) = 9.65, p = .02$ , SSQ<sub>Ocu</sub>,  $\chi^2(3) = 16.02, p = .001$ , and SSQ<sub>Dis</sub>,  $\chi^2(3) = 28.45, p < .001$ . Pairwise comparison revealed significant increases in PQ<sub>Invo</sub>,  $t(125) = 4.37, p < .001, r = .36$ , PQ<sub>ImrsAdpt</sub>,  $t(125) = 2.82, p = .03, r = .25$ , UTAUT<sub>HedMotv</sub>,  $t(126) = 2.56, p = .02, r = .22$ , SUS,  $t(126) = 2.93, p = .02, r = .25$ , and significant decreases in SSQ<sub>Ocu</sub>,  $t(126) = -3.71, p = .002, r = .31$ , and SSQ<sub>Dis</sub>,  $t(126) = -4.04, p < .001, r = .34$ , from S1 to S2. Similarly, there were significant increases in PQ<sub>Invo</sub>,  $t(125) = 3.98, p < .001, r = .34$ , PQ<sub>ImrsAdpt</sub>,  $t(125) = 2.43, p = .04, r = .25$ , and significant decreases in SSQ<sub>Ocu</sub>,  $t(126) = -3.13, p = .005, r = .26$  and SSQ<sub>Dis</sub>,  $t(126) = -4.35, p < .001, r = .36$ , from S1 to S4. Furthermore, there were significant decreases in UTAUT<sub>HedMotv</sub> from S2 to S3,  $t(126) = -2.80, p = .01, r = .24$ , and S2 to S4,  $t(126) = -3.46, p = .004, r = .22$ . There were no significant differences between individual sessions for PQ<sub>IntQual</sub> (all  $p > .05$ ).

**Discussion**

To our knowledge, this is the first randomized controlled study to investigate the psychological and psychophysiological effects of an HRV-BF intervention program supported by laboratory sessions in VR and at-home exercises using a mobile phone. Our results show that the four-week-long intervention significantly reduced chronic stress, anxiety, and mental fatigue and improved mindfulness and HRQoL (small effects). Specifically, all improvements in psychological traits post-intervention remained significant until follow-up, although chronic stress only significantly decreased



**Table 3** Effect of intervention on stress and stress-related symptoms: Interaction effect of *Time* × *Group* and planned contrasts following significant interaction effects

Variables	Time x Group		Pre – Post				Pre – Follow-up			
	$\chi^2$ (df)	p	b	t (df)	p	r	b	t (df)	p	r
PSS-10	6.23 (2)	<b>.04</b>	-1.37	-1.29 (152)	.20	.17	-2.65	-2.49 (152)	<b>.01</b>	.20
DASS-21 <sub>depression</sub>	2.66 (2)	.26								
DASS-21 <sub>anxiety</sub>	6.13 (2)	<b>.05</b>	-2.68	-2.16 (152)	<b>.03</b>	.17	-2.62	-2.11 (152)	<b>.04</b>	.17
DASS-21 <sub>stress</sub>	3.50 (2)	.17								
WHO-5	0.87 (2)	.64								
MAAS	11.24 (2)	<b>.00</b>	0.25	2.05 (152)	<b>.04</b>	.16	0.41	3.34 (152)	<b>.00</b>	.26
MFI-20 <sub>general fatigue</sub>	0.47 (2)	.79								
MFI-20 <sub>physical fatigue</sub>	2.71 (2)	.26								
MFI-20 <sub>mental fatigue</sub>	8.49 (2)	<b>.01</b>	-1.44	-2.41 (151)	<b>.02</b>	.19	-1.55	-2.61 (151)	<b>.01</b>	.21
MFI-20 <sub>reduced activity</sub>	3.96 (2)	.14								
MFI-20 <sub>reduced motivation</sub>	3.53 (2)	.17								
SF-36 <sub>general health</sub>	1.80 (2)	.41								
SF-36 <sub>mental health</sub>	4.11(2)	.13								
SF-36 <sub>bodily pain</sub>	2.10 (2)	.35								
SF-36 <sub>physical functioning</sub>	2.68 (2)	.26								
SF-36 <sub>role-emotional</sub>	7.40 (2)	<b>.02</b>	15.86	2.21 (152)	<b>.03</b>	.18	17.65	2.46 (152)	<b>.01</b>	.20
SF-36 <sub>role-physical</sub>	4.28 (2)	.11								
SF-36 <sub>social functioning</sub>	6.66 (2)	<b>.04</b>	15.86	2.21 (152)	<b>.03</b>	.18	17.65	2.46 (152)	<b>.02</b>	.20
SF-36 <sub>vitality</sub>	2.02 (2)	.37								
BPNSFS <sub>autonomy satisfaction</sub>	1.29 (2)	.45								
BPNSFS <sub>autonomy frustration</sub>	5.04 (2)	.08								
BPNSFS <sub>competence satisfaction</sub>	4.14 (2)	.13								
BPNSFS <sub>competence frustration</sub>	1.99 (2)	.37								
BPNSFS <sub>relatedness satisfaction</sub>	1.41 (2)	.49								
BPNSFS <sub>relatedness frustration</sub>	0.34 (2)	.84								
BPNSFS <sub>needs satisfaction</sub>	0.53 (2)	.77								
BPNSFS <sub>needs frustration</sub>	1.17 (2)	.56								
mean HR	6.00 (2)	<b>.05</b>	1.39	0.61 (146)	.54	.05	-4.00	-1.76 (146)	.08	.14
rMSSD	5.34 (2)	.06								
pNN50	7.68 (2)	<b>.02</b>	2.88	0.85 (146)	.40	.07	9.25	2.71 (146)	<b>.01</b>	.22
SDNN	5.30 (2)	.07								
HF	3.05 (2)	.22								
LF	9.94 (2)	<b>.01</b>	570.60	3.16 (146)	<b>.00</b>	.25	264.72	1.45 (146)	.15	.12
LF/HF	12.53 (2)	<b>.00</b>	0.86	1.91 (146)	.06	.15	-0.77	-1.69 (146)	.09	.14
SYS	0.57 (2)	.75								
DIA	5.86 (2)	.05								

Abbreviations: *PSS-10* Perceived Stress Scale, *DASS-21* Depression Anxiety Stress Scales, *WHO-5* World Health Organisation-Five Well-Being Index, *MAAS* Mindfulness Attention and Awareness Scale, *MFI-20* Multidimensional Fatigue Inventory, *SF-36* Short Form Health Survey, *BPNSFS* Basic Psychological Need Satisfaction and Frustration Scale, *HR* Heart rate, *rMSSD* Root Mean Square of Successive Differences, *pNN50* Percentage of successive normal-to-normal heartbeat intervals differing by > 50 ms, *SDNN* Standard deviation of normal-to-normal heartbeat intervals, *HF* High-frequency power, *LF* Low frequency power, *SYS* Systolic blood pressure, *DIA* Diastolic blood pressure, *r* effect size Pearson's *r*

Significant *p*-values are marked in bold (also includes significant *p*-values that have been rounded up to .05)

when comparing levels at pre-intervention to follow-up. The intervention was also able to significantly increase measures of HRV, specifically pNN50, LF, and LF/HF

ratio compared to the WLC. Differences in LF and LF/HF ratio between groups did not last until follow-up, while differences in pNN50 were only observed when

comparing changes from pre-intervention to follow-up. Although changes in resting HR were significantly different in the two groups over time, planned contrasts revealed no significant differences between measurements. There was only a visual trend for a decrease in resting HR from pre-intervention to follow-up in the INT compared to the WLC. No differences in terms of resting BP or psychological and psychobiological stress reactivity, however, were found between the two groups. As for UX measures, results show that feeling autonomously motivated towards using the technology and the intervention significantly decreased, while perceived competence significantly increased over time. There was also a considerable rise from the first to the last of four sessions in both involvement and immersion. Finally, hedonic motivation was highest in the second session before progressively reverting to initial values.

Mean baseline values of psychological outcomes were mostly in line with other samples [53, 99, 112, 113]. Specifically, long-term stress levels, as indicated by PSS-10 and DASS-21<sub>stress</sub>, were not elevated [91, 92, 114]. The sample, however, had mild depressive symptoms according to the DASS-21<sub>depression</sub> categories [92] and “fair” psychological well-being as defined by the WHO-5 benchmarks [115]. Baseline values of physiological outcomes such as mean HR and LF can be situated within the range of short-term HRV norms [116–118]. Concerning SYS, baseline values were slightly elevated in both groups for the age group of our sample, whereas baseline DIA was comparable to guidelines and norm values [119, 120].

The findings indicate that using VR to support the delivery of HRV-BF improves a range of indicators of stress and stress-related symptoms to the same extent as HRV-BF interventions delivered on two-dimensional screens (see [21–23]). We observed small improvements across a range of psychological traits (i.e., chronic stress, anxiety, mental fatigue, mindfulness, HRQoL) and small changes in measures of cardiac activity (i.e., pNN50, LF, and LF/HF ratio) in the intervention compared to the WLC. This pattern and extent of changes in primary and secondary outcomes are comparable to those found in Lehrer et al.’s [22] meta-analysis. Since effect sizes are modest and similar in size across measures, Lehrer et al. [22] conclude that HRV-BF may be useful for targeted

and non-targeted measures and serve as a complementary treatment method.

Due to the lack of an active control group, improvements in psychological outcomes in the intervention group due to potential placebo effects could have been caused by participants’ expectation to improve and higher motivation compared to the WLC due to the multifaceted, novel, and engaging nature of the intervention. In addition, experimenter demands effects could have played a role, such that participants consciously or unconsciously aimed to produce effects believed to be expected by the experimenters. While the absence of an active control group prohibits us from a definite conclusion on effect sizes, there is evidence provided by the data suggesting that placebo or experimenter demands effects may be somewhat negligible.

First, there is the observation of significant differences between groups in some physiological measures, especially the significant increase in resting LF, the prime target measure of HRV-BF, from pre- to post-intervention in the intervention group. The increases in the main target parameter of HRV-BF, the LF band, presumably reflect increased baroreflex activity during resting conditions [29, 51, 121]. There were also increases in pNN50 from pre- to post-intervention and follow-up, and the visual trends of an attenuated endocrine stress reactivity is a pattern observed in other SMI studies [122–124].

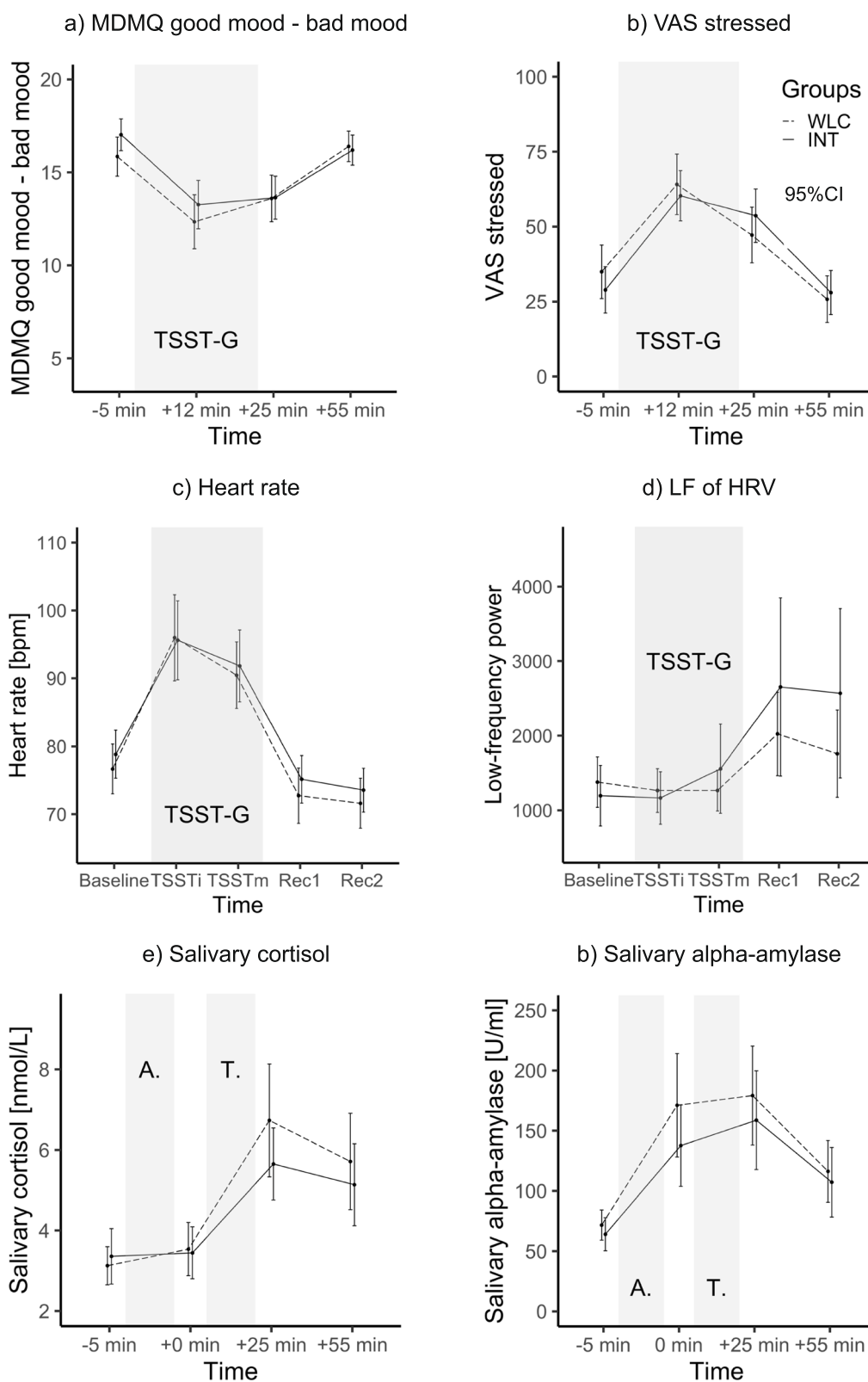
Second, self-reported chronic stress decreased not until follow-up, although the expectation of experiencing stress-reducing effects would have been most salient to participants immediately after the intervention. Moreover, participants reported decreases in autonomous regulation and increases in controlled regulations toward using the technologies and intervention over time. This pattern of results challenges the notion of a pervasive positive bias in evaluating intervention features and outcomes.

Third, there is evidence provided by research on HRV-BF for stress management with active control groups showing that HRV-BF is indeed effective in reducing a variety of symptoms. De Bruin [125] and van der Zwan [126] both report on the same study but with different outcomes. The randomized controlled trial compared the effects of HRV-BF ( $n = 25$ ) to mindfulness meditation ( $n = 27$ ) and physical

(See figure on next page.)

**Fig. 5** Group means of repeated psychological and biological measures of stress reactivity. Error bars indicate 95% confidence intervals.

Abbreviations: INT = Intervention group; WLC = Wait-list control group; TSST-G = Trier Social Stress Test for Groups; TSSTi = Mock job interview; TSSTm = Mental arithmetic task; Rec1 = Recovery phase 1; Rec2 = Recovery phase 2; A. = Anticipation; T. = Trier Social Stress Test for Groups; MDMQ = Multidimensional Mood Questionnaire; VAS = Visual Analog Scale; bpm = beats per minute; LF of HRV = Low-frequency power of heart rate variability; nmol/L = Nanomole per liter; U/ml = Units per milliliter. Error bars indicate 95% confidence intervals



**Fig. 5** (See legend on previous page.)

activity ( $n = 23$ ) for stress reduction in a stressed sample of young adults. The interventions included daily exercises for five weeks, similar to our intervention in terms of effort and duration. The articles report significant improvements in measures of stress, anxiety, worrying, mindful awareness, psychological well-being, and self-compassion, among others, with small to moderate effect sizes. These improvements were observed from pre- to post-intervention and pre-intervention to follow-up six weeks post-intervention in all groups. There were no significant between-group differences, however, concluding that HRV-BF is not more but equally effective for stress and anxiety as the other methods [125, 126]. This pattern is also backed by meta-analytic computations by Lehrer et al. [22]. Compared to active controls (e.g., EEG biofeedback, physical exercise, cognitive behavioral therapy, skill training, mindfulness meditation, monitoring HRV, progressive muscle relaxation), HRV-BF does not lead to significantly better outcomes. In contrast, compared to inactive controls (e.g., WLC, no treatment, sitting quietly, watching a video), HRV-BF achieves significant small to moderate effect sizes [22].

In conclusion, the results of the present study support the notion that HRV-BF supported by VR and mobile technology is also able to reduce stress and stress-related symptoms and modulate emotion-regulating networks of the brain through the baroreflex system (bottom-up processes) [22, 51] and that it is also able to train prefrontal regions of the brain to exert regulatory control over areas that are involved in perceiving and experiencing emotions (top-down processes) [33, 37, 127]. The observation that dimensions of HRQoL, such as role-emotional and social functioning, improved might be explained by the positive effects of vagal activation on self-regulatory processes in the brain. According to Porges [29], healthy vagal functioning inhibits sympathetic outflow to the heart and allows humans to self-regulate, feel safe, and show adaptive social behavior.

We assume that the repeated, near-to-daily RF breathing mainly drove the intervention effects during the laboratory sessions in VR and the mobile-supported exercises. According to CAMIL [62] and empirical findings, different technologies can be particularly conducive to teaching and learning a method or a skill, but they are not the main drivers of effects. Our previous work [53] showed that standardized slow and paced breathing is the main contributor to short-term increases in LF and CR, while RF breathing coupled with biofeedback is able to significantly amplify this effect. Similarly, both HMDs and two-dimensional screens can be used for both breathing types, but using an HMD was able to significantly amplify the effects on CR. Nevertheless, the

individual contributions of VR-supported HRV-BF and mobile-supported breathing at RF cannot be discerned.

Some of the assessed psychological and psychobiological outcomes did not differ between conditions. For example, although HRV-BF has been shown to improve BP control modulated by the baroreflex [85] and recent meta-analyses confirming that HRV-BF interventions significantly reduce SYS and DIA [22, 26], we did not observe any differences between the INT and the WLC. However, in our study, BP levels also decreased in the control group from *Pre* to *Post* intervention. Moreover, although research assistants interacted only minimally with participants during the BP assessments, we cannot rule out a white coat effect (not only, but especially during the *Pre* measurement) which might also explain the slightly elevated levels of SYS throughout the study in this otherwise healthy sample. In addition, the VR-supported HRV-BF intervention at hand did not significantly affect psychological and psychobiological stress reactivity, except for the visual trends in LF, sCort, and sAA. Changes in mean values of LF seem to suggest greater power in the LF band from peak reactivity (i.e., mental arithmetic task) throughout recovery in the INT compared to the WLC, perhaps reflecting a more adaptive recovery of the autonomous nervous system (ANS). Notably, the values during recovery are higher than at baseline in the INT, which might suggest that some participants in the INT actively used slow and paced breathing to recover from the psychosocial stress test. Similarly, the trends of changes in mean levels of sCort and sAA could reflect a less pronounced stress response of the HPA axis and the sympathetic nervous system in the INT compared to the WLC. Participants in the INT might have also been particularly acclimatized to the laboratory setting, familiar with the research assistants present in the room and the participants in their slots. A sense of familiarity associated with increased resilience [128] in the INT might have led to a less pronounced stress response. Alternatively, one could also have expected that improved resting parameters are accompanied by stronger stress reactivity, possibly explained by the potential of showing a greater net reduction in vagal activity and a more solid increase in sympathetic activity in response to challenging demands [129, 130]. Nonetheless, the TSST-G in this study was able to induce significant psychological and psychobiological stress responses in both groups, although the increases in sCort were not as high as in other studies employing the same protocol (e.g., [89, 113]). According to a meta-analysis [131], having three members, instead of one or two, on the panel of judges and using the number 13 instead of 17 can increase the cortisol response. The time at which this study was conducted (May to June 2022) might also have



contributed to the comparatively small improvements in measures in the INT and the slight worsening of some outcomes in the WLC. During this period, students typically transition into the end-of-semester examination phase, preparing for or already taking exams. Study participation might thus have attenuated improvements in the INT or even contributed to stress and stress-related symptoms in both groups. Moreover, the daily workload in the INT and having to come to the lab once a week over a month might have further increased pressure and stress in this group, potentially adding to diminished effectiveness. Log data of the mobile application also suggest that adherence to the exercises was high overall but did decrease slightly from the start to the end of the program, potentially contributing to the attenuated effectiveness.

The decrease in autonomous regulation and the increase in controlled regulation towards the adoption and use of the technology from pre- to post-intervention also support this explanation. Having to do the exercises with the mobile application every single day without any BF might have left participants feeling less in control and without agency, and therefore stressed and frustrated. The observation that levels of chronic stress only showed improvements at follow-up once the intervention was really over might also back this explanation. Being autonomously motivated to use technology and participate in health interventions is crucial to achieving better health and well-being outcomes as well as to experiencing overall less stress and negative affect (e.g., [73, 74, 132–136]). The fact that participants were remunerated for their participation may have also contributed to feeling more controlled and less autonomous. Years of research have shown that tangible rewards such as money or trophies feel controlling and thereby undermine intrinsic motivation [137]. On the contrary, rewards in the form of positive feedback that carry information about performance, such as a cumulative score during HRV-BF, for instance, have been shown to increase intrinsic motivation [137]. However, this is the case only when individuals' perceived competence is accompanied by some degree of perceived autonomy, an internal locus of control [137]. Autonomy is considered to be an important principle from the perspective of biomedical ethics [138], which should be safeguarded and fostered throughout the design of any technology-supported stress management intervention [139]. Participants' free text feedback indicated that they wished for the ability to personalize the virtual environment and add BF to the mobile application. Moreover, considering that 10 participants who were allocated to the INT did not show up for the first session (compared to two participants in the WLC), the schedule of the intervention

program might have been viewed as too intense, stressful, and inflexible already in advance. Regardless of the decreases in autonomy, the increase in perceived competence over time shows that the system did become easier to use, and participants felt more effective at using it and less challenged over time. This might entail that participants got more confident in using the technology and devices and in navigating the virtual environment and mastering the breathing exercises and the BF task. Experiencing increased competence in a task can ultimately benefit intrinsic motivation [71].

The remaining UX assessment of the intervention during the four lab sessions of the INT yielded overall positive results among all dimensions. The mean values of UX measures were in the upper third of the scale ranges, except for interface quality, involvement, and sensor fidelity, which were all in the upper half. Due to the lack of long-term studies, we can only compare our results to previous research using either the mean values of UX measures from the first session or the mean values across all four sessions. Here, the means of both the first session and across sessions are similar or higher to values reported in previous research using VR for BF [65, 68, 69, 140–142]. Furthermore, we found that the UX measures of the first session were overall similar or higher than values measured for HMDs in our previous work [53], indicating the improvements made to the VE had the intended effect. CAMIL proposes that using HMDs for VR experiences can improve learning outcomes due to higher values of affective and cognitive factors (e.g., motivation, embodiment, and self-efficacy) [62]. The fact that the UX values stayed high throughout all four sessions thus indicates that the use of HMDs could be beneficial in comparison to a screen not only in a single session but also in the long run.

As for changes in UX measures between sessions, we found significant increases from the first to the second session, after which the measures either stayed on the same level (i.e., involvement and immersion) or returned to the initial values of the first session (i.e., motivation and usability). These results indicate that the participants did not experience a novelty effect surrounding our system (i.e., initial excitement towards new technologies that wears off with repeated use). The absence of a novelty effect is important since sustained high UX is expected to positively affect the intervention outcome [62]. Furthermore, adherence to an intervention program is reliant on sustained motivation [75]. The lack of further improvements in UX measures after the second session could also be related to the corresponding changes in the interface designed to support the shift from external to internal feedback (i.e., removing the pacer and adding real-time feedback).

Although participants seemed to adapt to these changes, as indicated by the moderate increases in perceived competence, this pattern of results might also suggest that participants would require slightly more time to adapt to the changes or that BF should be included in the mobile application for homework. This is further supported by participants' free text feedback, indicating that it became more difficult to perform the training in later sessions (i.e., the third and fourth sessions). Finally, we observed overall low levels of simulator sickness symptoms and a small but significant reduction (medium effect size) of symptoms over time, indicating that the technology used for the intervention did not hinder participants' user experience or knowledge acquisition.

### Limitations, future research, and implications

This study comes with a series of limitations. First, a major limitation of this study is that it did not include an active control group with comparable expectations of improvement (e.g., sham biofeedback, EEG feedback, cognitive behavioral therapy, mindfulness meditation, or physical exercise), which could have controlled for potential placebo effects in the intervention group. Second, although some measures worsened in the WLC over time, some psychological and psychophysiological outcomes indicate improvements (e.g., HRQoL, physical fatigue, DIA, and SYS). Even more pronounced positive effects in control groups relative to treatment groups have been reported in other HRV-BF intervention studies [143, 144]. Possible treatment diffusion [145] may explain these positive effects. All participants were informed about the general nature of the study (e.g., HRV-BF in VR, breathing exercises) before giving informed consent. Therefore, participants in the WLC may have felt inspired to practice breathing exercises or other stress management skills during the time in which the other participants received the intervention. Third, the sample consisted of healthy young students, limiting the generalisability to older, highly stressed individuals and clinical populations. Offering the intervention to populations with elevated stress levels might have led to more and greater improvements across measures. Fourth, the study only included one follow-up assessment four weeks post-intervention. The lack of later follow-up assessments at three, six, or perhaps twelve months do not allow conclusions on the long-lasting and sustained effects of this intervention beyond a month. Lastly, significant changes in primary and secondary outcomes warrant cautious interpretation and conclusion of implications considering the small effect sizes detected with the sample size at hand.

Based on the results of this study, we also see a series of avenues for future research and practical implications.

First, introducing different and greater degrees of personalization of the virtual environment and BF might increase participants' autonomy. In the context of our VR-supported HRV-BF protocol, autonomy might have been higher by integrating BF in the mobile application and adapting the BF according to individual progress. Using BF during homework exercises with the mobile application could be tested in combination with the VR application in the lab or perhaps provide an entirely autonomous BF system outside of a lab setting at users' disposal. The more control over time, day, place, and circumstances participants are given, perhaps the higher their perceived autonomy and adherence and, thus, the beneficial effects of the intervention. Second, the intervention program could be tested over a longer period, as it could further improve the effects observed in the current study and lead to more pronounced differences compared to a control group. Additionally, it might be worth investigating whether different durations and frequencies of homework exercises impact the effectiveness of the intervention program. Fourth, as the technology is ever-evolving and improving, further research with both more advanced but also more simplified VR devices could be tested. This could include comparing HRV-BF in VR to standardized paced breathing in VR as an active control group. Alternatively, affordable VR devices could be handed out to participants, moving the intervention from the laboratory to the field. In such a field experiment, a standalone VR-supported HRV-BF setup could be compared to an entirely mobile-supported HRV-BF. This would reduce the complexity of the interventions and help isolate the effects of the different technologies. Fifth, the intervention program could be tested with adolescent populations, which have a high prevalence of stress and mental health issues [146] and might be especially affine and interested in using new technologies. Using new technologies might foster the education of young people about stress and teach them strategies to cope with stressful situations and challenging life events. As shown in our previous work [53], VR seems especially effective in practicing HRV-BF or paced breathing exercises and thereby teaching stress management skills. Lastly, the current findings need to be replicated and proven to be robust in randomized clinical trials in various settings and populations with adequate controls and larger sample sizes before it is offered as treatment. Assuming effectiveness, and higher degrees of independence and automatization of the psychoeducational content, system, and data processing, we can envision the practical use of such a VR-supported HRV-BF intervention program in the near future. This could include the private use or the use at the workplace as part of a wider stress management offer, the use in clinical settings such as psychiatric clinics, nursing homes, rehabilitation, or

pain management centers to support the treatment of stress-related disorders and the management of chronic diseases. To facilitate this, full integration of the BF system and psychoeducational into a stand-alone application, easy connectivity to wearable devices, and guidance through a week-long program must be guaranteed. In a clinical setting, such VR-supported HRV-BF standalone could be offered as a complementary intervention alongside standard treatments. Such a standalone offer would not require much time and effort from healthcare personnel to introduce the intervention to patients. It could be easily integrated into a patient's treatment plan, both for in- and outpatients. Patients could use such an application autonomously in their own time, even beyond the hospitalization period supported by the mobile application. This program could be offered to individuals wanting to prevent or alleviate chronic stress and improve their stress management and emotion regulation skills.

## Conclusion

The VR and mobile-supported HRV-BF intervention program presented in this work improved various stress indicators and stress-related symptoms compared to a WLC group. The findings of this study provide preliminary evidence for the suitability of using VR technology for laboratory sessions of HRV-BF training. The sizes of beneficial effects are comparable to the effects reported in studies on HRV-BF interventions delivering laboratory sessions on two-dimensional screens. Considering the decreased autonomous regulation regarding the technology and intervention over time, an emphasis should be placed on autonomy-supportive interventions that include higher degrees of flexibility and adaptability, which should ultimately benefit long-term health and well-being. The good to very good UX ratings support the future use of HMDs for HRV-BF and continued investigations in this field. The absence of a novelty effect regarding technology adoption further supports using such devices for an HRV-BF intervention program, as a high UX throughout the program can lead to higher adherence and yield more beneficial outcomes. Concerning future research, the long-term psychological and psychobiological benefits of VR delivery compared to different screen modalities remain to be studied. The suitability and effectiveness of a VR-supported HRV-BF intervention for highly stressed individuals and clinical samples also require further investigation. Future studies should address potential placebo effects by comparing VR-supported HRV-BF to an active control group, equally expecting to improve. All in all, this work contributes to a growing field of research leveraging new and increasingly affordable digital technologies, including wearables, VR, and smartphones, to promote health and transform healthcare.

## Abbreviations

ACTA	Autonomy and Competence in Technology Adoption Scale Questionnaire
ACTA <sub>ARS</sub>	Autonomy Regulation Score of the Autonomy and Competence in Technology Adoption Scale Questionnaire
ACTA <sub>CRS</sub>	Controlled Regulation Score of the Autonomy and Competence in Technology Adoption Scale Questionnaire
ACTA <sub>RAI</sub>	Perceived Competence Score of the Autonomy and Competence in Technology Adoption Scale Questionnaire
ACTA <sub>PCS</sub>	Relative Autonomy Score of the Autonomy and Competence in Technology Adoption Scale Questionnaire
ANS	Autonomic nervous system
BP	Blood pressure
BPNSFS	Basic Psychological Needs and Frustration Scale
BF	Biofeedback
CAMIL	Cognitive Affective Model of Immersive Learning
COVID-19	Coronavirus Disease of 2019
CR	Coherence ratio
DASS-21	Depression, Anxiety and Stress Scales 21
DASS-21 <sub>anxiety</sub>	Anxiety Scale of the Depression, Anxiety, and Stress Scales
DASS-21 <sub>depression</sub>	Depression Scale of the Depression, Anxiety, and Stress Scales
DASS-21 <sub>stress</sub>	Stress Scale of the Depression, Anxiety, and Stress Scales
DIA	Diastolic blood pressure
FSS	Flow Short Scale
Follow-up	Measurements at follow-up
HF	High-frequency power of heart rate variability
HMD	Head-mounted display
HR	Heart rate
HRQoL	Health-related quality of life
HRV	Heart rate variability
HRV-BF	Heart rate variability biofeedback
HPA axis	Hypothalamic-pituitary-adrenal axis
INT	Intervention
LF	Low-frequency power of heart rate variability
MAAS	The Mindful Attention Awareness Scale
MDMQ	Multidimensional Mood Questionnaire
MDMQ <sub>mood</sub>	Good mood-bad mood scale of the Multidimensional Mood Questionnaire
MDMQ <sub>calmness</sub>	Calmness-nervousness scale of the Multidimensional Mood Questionnaire
MDMQ <sub>wakefulness</sub>	Wakefulness-sleepiness scale of the Multidimensional Mood Questionnaire
MFI-20	Mental Fatigue Inventory 20
nmol/L	Nanomole per liter
OSF	Open Science Framework
PQ	Presence Questionnaire
PQ <sub>ImrsAdpt</sub>	Immersion Adaption Scale of the Presence Questionnaire
PQ <sub>Invo</sub>	Involvement Scale of the Presence Questionnaire
PQ <sub>IntQual</sub>	Interface Quality Scale of the Presence Questionnaire
PQ <sub>SensFi</sub>	Sensor Fidelity Scale of the Presence Questionnaire
Post	Measurements post-intervention
Pre	Measurements pre-intervention
pNNS50	Percentage of successive normal-to-normal intervals differing by > 50 ms
PSS-10	Perceived Stress Scale 10
rMSSD	Root mean square of successive differences between normal heartbeats
RF	Resonance Frequency
RRI	R-R interval
SDNN	Standard deviation of normal-to-normal heartbeat intervals
S1	Session 1
S2	Session 2

S3	Session 3
S4	Session 4
sAA	Salivary alpha-amylase
sCort	Salivary cortisol
SF-36	Short Form Survey 36
SSQ	Simulator Sickness Questionnaire
SUS	System Usability Scale
SSQ <sub>Dis</sub>	Disorientation of the Simulator Sickness Questionnaire
SSQ <sub>Ocu</sub>	Oculomotor Scale of the Simulator Sickness Questionnaire
SSQ <sub>Nas</sub>	Nausea Scale of the Simulator Sickness Questionnaire
SYS	Systolic Blood Pressure
TSST-G	Trier Social Stress Test for Groups
TSST <sub>BL</sub>	Cardiac measurement before the Trier Social Stress Test
TSST <sub>I</sub>	Mock job interview phase of the Trier Social Stress Test
TSST <sub>M</sub>	Mental arithmetic task of the Trier Social Stress Test
TSST <sub>REC1</sub>	Cardiac measurement during the first phase of recovery after the Trier Social Stress Test
TSST <sub>REC2</sub>	Cardiac measurement during the second phase of recovery after the Trier Social Stress Test
U/mL	Unit per milliliter
UTAUT	Unified Theory of Acceptance and Use of Technology Questionnaire
UTAUT <sub>HedMotv</sub>	Hedonic Motivation of the Unified Theory of Acceptance and Use of Technology Questionnaire
UTAUT <sub>BhvInt</sub>	Behavioral Intention of the Unified Theory of Acceptance and Use of Technology Questionnaire
UTAUT <sub>EffExp</sub>	Effort Expectancy of the Unified Theory of Acceptance and Use of Technology Questionnaire
UTAUT <sub>FacilCond</sub>	Facilitating Conditions of the Unified Theory of Acceptance and Use of Technology Questionnaire
UX	User experience
U/mL	Units per milliliter
VAS	Visual Analog Scale
VE	Virtual environment
VR	Virtual reality
WLC	Wait-list control

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s44247-023-00042-z>.

**Additional file 1.** Appendix containing Tables A1-A4.

## Acknowledgements

We thank Dr. Stefan Wehrli of ETH Zurich's Decision Science Laboratory, Dr. Erika Meins, Dr. Sebastian Tillmanns, Michaela Benk, Caterina Bérubé, and Adam Fehr for their assistance in the realization of the study, and Dipl.-Psych. Dr. Nadine Skoluda of the biochemical laboratory of the Faculty of Psychology at the University of Vienna and Dr. Arran Davis of the University of Oxford for their work and support with sample/data analysis.

## Authors' contributions

J.J.K.: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. R.P.W.: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. M.N.: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – review & editing, Visualization, Project administration, Funding acquisition. A.F.: Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition. V.R.S.: Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. R.L.M.: Conceptualization, Resources, Writing – review & editing, Funding acquisition. C.H.: Writing – review & editing, Funding acquisition. U.M.N.: Methodology, Resources, Writing – review & editing, Supervision. F.v.W.: Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Funding acquisition.

## Funding

Open access funding provided by Swiss Federal Institute of Technology Zurich. This study is part of a larger project and supported by an ETH Research Grant, Switzerland (ETH-09 19-2). The funding body did not play a role in the design of the study, in the collection, analysis, and interpretation of data, or in writing the manuscript.

## Availability of data and materials

The datasets supporting the conclusions of this article are available in the Open Science Framework (OSF) repository in <https://osf.io/ge8h9/>.

## Declarations

### Ethics approval and consent to participate

Participants gave written informed consent before participation. The study was approved by the Ethics Commission of ETH Zurich (2022-N-35) and conducted in accordance with the Declaration of Helsinki. The study was pre-registered on OSF registries [147].

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Mobililar Lab for Analytics at ETH Zurich, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland. <sup>2</sup>Chair of Technology Marketing, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland. <sup>3</sup>Faculty of Society & Design, Bond University, Gold Coast, Australia. <sup>4</sup>Centre for Stress-Related Disorders, Clinica Holistica Engiadina, Susch, Switzerland. <sup>5</sup>Chair of Clinical Psychology and Psychotherapy, Department of Psychology, University of Zurich, Zurich, Switzerland. <sup>6</sup>Chair of Cognitive Science, Department of Humanities, Social and Political Sciences, ETH Zurich, Zurich, Switzerland. <sup>7</sup>Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore, Singapore. <sup>8</sup>Chair of Clinical Psychology of Adulthood, Department for Clinical and Health Psychology, University of Vienna, Vienna, Austria. <sup>9</sup>Research Platform The Stress of Life (SOLE) – Processes and Mechanisms underlying Everyday Life Stress, University of Vienna, Vienna, Austria.

Received: 27 April 2023 Accepted: 5 September 2023

Published online: 13 October 2023

## References

- Sparks K, Faragher B, Cooper CL. Well-being and occupational health in the 21st century workplace. *J Occup Organ Psychol*. 2001;74(4):489–509. <https://doi.org/10.1348/096317901167497>.
- Ulshöfer C, Jensen R, Galliker S, Igic I, Elfering A, Semmer N. Job-Stress-Index 2022. Monitoring von Kennzahlen zum Stress bei Erwerbstätigen in der Schweiz. *Gesundheitsförderung Schweiz*; 2022. <https://gesundheit.sfoerderung.ch/news/gesundheitsfoerderung-und-praevention-allgemein/gesundheits-bei-der-arbeit-emotionale>. Accessed 04 Apr 2023.
- de Quervain, D, Aerni, A, Amin, E, Benz, D, Coynel, D, Gerhards, C, Zuber, P. The Swiss Corona Stress Study; 2020. OSF Preprints. <https://doi.org/10.31219/osf.io/jqw6a>.
- Galliker S, Igic I, Elfering A, Semmer-Tschan N, Brunner B, Wieser S. Job-Stress-Index 2018. Kennzahlen zum Stress bei Erwerbstätigen in der Schweiz. *Gesundheitsförderung Schweiz*; 2018. <https://gesundheitsfoerderung.ch/ueber-uns/medien/medienmitteilungen/artikel/job-stress-index-2018-jede-vierte-erwerbstaetige-person-hat-stress.html>. Accessed 04 Apr 2023.
- Grebner S, Elfering A, Semmer NK. The success resource model of job stress. In: *New developments in theoretical and conceptual approaches to job stress*, vol 8. Bingley: Emerald Group Publishing Limited; 2010. p. 61–108. [https://doi.org/10.1108/S1479-3555\(2010\)0000008005](https://doi.org/10.1108/S1479-3555(2010)0000008005).
- Perriard DRJ. Die Kosten des Stresses in der Schweiz. 2003. [https://www.seco.admin.ch/seco/de/home/Publikationen\\_Dienstleistungen/Publikatio](https://www.seco.admin.ch/seco/de/home/Publikationen_Dienstleistungen/Publikatio)



- nen\_und\_Formulare/Arbeit/Arbeitsbedingungen/Studien\_und\_Berichte/die-kosten-des-stresses-in-der-schweiz.html. Accessed 04 Apr 2023.
7. Lazarus RS, DeLongis A. Psychological stress and coping in aging. *Am Psychol*. 1983;38(3):245. <https://doi.org/10.1037//0003-066x.38.3.245>.
  8. Bakker AB, Demerouti E. The job demands-resources model: state of the art. *J Manag Psychol*. 2007;22(3):309–28. <https://doi.org/10.1108/02683940710733115f>.
  9. Chrousos GP. Stress and disorders of the stress system. *Nat Rev Endocrinol*. 2009;5(7):374–81. <https://doi.org/10.1038/nrendo.2009.106>.
  10. McEwen BS. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiol Rev*. 2007;87(3):873–904. <https://doi.org/10.1152/physrev.00041.2006>.
  11. Grados JL. Prevalence and prognosis of stress disorders: a review of the epidemiologic literature. *Clin Epidemiol*. 2017;9:251–60.
  12. Finlay-Jones R, Brown GW. Types of stressful life event and the onset of anxiety and depressive disorders. *Psychol Med*. 1981;11(4):803–15. <https://doi.org/10.1017/s0033291700041301>.
  13. Hammen C. Stress and depression. *Annu Rev Clin Psychol*. 2005;1(1):293–319. <https://doi.org/10.1146/annurev.clinpsy.1.102803.143938>.
  14. Preckel D, von Känel R, Kudiella B, Fischer J. Overcommitment to work is associated with vital exhaustion. *Int Arch Occup Environ Health*. 2005;78:117–22. <https://doi.org/10.1007/s00420-004-0572-8>.
  15. Kivimäki M, Kawachi I. Work stress as a risk factor for cardiovascular disease. *Curr Cardiol Rep*. 2015;17(9):74.
  16. Ames SC, Jones GN, Howe JT, Brantley PJ. A prospective study of the impact of stress on quality of life: An investigation of low-income individuals with hypertension. *Ann Behav Med*. 2001;23(2):112–9. [https://doi.org/10.1207/S15324796ABM2302\\_5](https://doi.org/10.1207/S15324796ABM2302_5).
  17. Epel ES, Crosswell AD, Mayer SE, Prather AA, Slavich GM, Puterman E, et al. More than a feeling: A unified view of stress measurement for population science. *Front Neuroendocrinol*. 2018;49:146–69. <https://doi.org/10.1016/j.yfrne.2018.03.001>.
  18. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. *J Behav Med*. 2019;42:67–83. <https://doi.org/10.1007/s10865-018-9966-z>.
  19. Kennedy L, Parker SH. Biofeedback as a stress management tool: a systematic review. *Cogn Tech Work*. 2019;21(2):161–90. <https://doi.org/10.1007/s10111-018-0487-x>.
  20. Association for Applied Psychophysiology and Biofeedback. About BioFeedback. 2023. [https://aapb.org/About\\_BioFeedback](https://aapb.org/About_BioFeedback). Accessed 18 July 2023.
  21. Goessl VC, Curtiss JE, Hofmann SG. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. *Psychol Med*. 2017;47(15):2578–86. <https://doi.org/10.1017/S0033291717001003>.
  22. 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*. 2020;45(3):109–29. <https://doi.org/10.1007/s10484-020-09466-z>.
  23. Pizzoli SF, Marzorati C, Gatti D, Monzani D, Mazzocco K, Pravettoni G. A meta-analysis on heart rate variability biofeedback and depressive symptoms. *Sci Rep*. 2021;11(1):6650. <https://doi.org/10.1038/s41598-021-86149-7>.
  24. Fournié C, Chouchou F, Dalleau G, Caderby T, Cabrera Q, Verkindt C. Heart rate variability biofeedback in chronic disease management: a systematic review. *Complement Ther Med*. 2021;60:102750. <https://doi.org/10.1016/j.ctim.2021.102750>.
  25. De Witte I N A J Buyck, Van Daele T. Combining biofeedback with stress management interventions: a systematic review of physiological and psychological effects. *Appl Psychophysiol Biofeedback*. 2019;44:71–82. <https://doi.org/10.1007/s10484-018-09427-7>.
  26. Costa Vital JE, de Moraes Nunes A, Souza de Albuquerque Cacique New York B, Araujo de Sousa BD, Nascimento MF, Formiga MF, et al. Biofeedback therapeutic effects on blood pressure levels in hypertensive individuals: a systematic review and meta-analysis. *Complement Ther Clin Pract*. 2021;44:101420. <https://doi.org/10.1016/j.ctcp.2021.101420>.
  27. Berntson GG, Quigley KS, Norman GJ, Lozano DL. Cardiovascular psychophysiology. In: *Handbook of psychophysiology*. 4th ed. New York: Cambridge University Press; 2017. p. 183–216. <https://doi.org/10.1017/9781107415782.009>.
  28. of the European Society of Cardiology the North American Society of Pacing Electrophysiology TF. Heart rate variability. *Circulation*. 1996;93(5):1043–1065. <https://doi.org/10.1161/01.CIR.93.5.1043>.
  29. Porges SW. The polyvagal perspective. *Biol Psychol*. 2007;74(2):116–43. <https://doi.org/10.1016/j.biopsycho.2006.06.009>.
  30. Heiss S, Vaschillo B, Vaschillo EG, Timko CA, Hormes JM. Heart rate variability as a biobehavioral marker of diverse psychopathologies: a review and argument for an “ideal range”. *Neurosci Biobehav Rev*. 2021;121:44–55. <https://doi.org/10.1016/j.neubiorev.2020.12.004>.
  31. Gidron Y, Deschepper R, De Couck M, Thayer JF, Velkeniers B. The vagus nerve can predict and possibly modulate non-communicable chronic diseases: introducing a neuroimmunological paradigm to public health. *J Clin Med*. 2018;7(10). <https://doi.org/10.3390/jcm7100371>.
  32. Vrijkotte TGM, van Doornen LJP, de Geus EJC. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*. 2000;35(4):880–6. <https://doi.org/10.1161/01.HYP.35.4.880>.
  33. Thayer JF, Lane RD. A model of neurovisceral integration in emotion regulation and dysregulation. *J Affect Disord*. 2000;61(3):201–16. [https://doi.org/10.1016/s0165-0327\(00\)00338-4](https://doi.org/10.1016/s0165-0327(00)00338-4).
  34. Dell’Acqua C, Dal Bò E, Messerotti Benvenuti S, Palomba D. Reduced heart rate variability is associated with vulnerability to depression. *J Affect Disord Rep*. 2020;1:100006. <https://doi.org/10.1016/j.jadr.2020.100006>.
  35. MacKinnon S, Gevirtz R, Rollin M, Brown M. Utilizing heartbeat evoked potentials to identify cardiac regulation of vagal afferents during emotion and resonant breathing. *Appl Psychophysiol Biofeedback*. 2013;38:241–55. <https://doi.org/10.1007/s10484-013-9226-5>.
  36. McCraty R, Shaffer F. Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. *Global Adv Health Med*. 2015;4(1):46–61. <https://doi.org/10.7453/gahmj.2014.073>.
  37. Schumann A, Cruz Fdl, Köhler S, Brotte L, Bär KJ. The influence of heart rate variability biofeedback on cardiac regulation and functional brain connectivity. *Front Neurosci*. 2021;15:691988. <https://doi.org/10.3389/fnins.2021.691988>.
  38. Lehrer PM, Vaschillo E, Vaschillo B. Resonant frequency biofeedback training to increase cardiac variability: rationale and manual for training. *Appl Psychophysiol Biofeedback*. 2000;25:177–91. <https://doi.org/10.1023/A:1009554825745>.
  39. Vaschillo EG, Vaschillo B, Lehrer PM. Characteristics of resonance in heart rate variability stimulated by biofeedback. *Appl Psychophysiol Biofeedback*. 2006;31(2):129–42. <https://doi.org/10.1007/s10484-006-9009-3>.
  40. Lehrer PM, Gevirtz R. Heart rate variability biofeedback: how and why does it work? *Front Psychol*. 2014;5. <https://doi.org/10.3389/fpsyg.2014.00756>.
  41. Hallman DM, Olsson EMG, Schéelle Bv, Melin L, Lyskov E. Effects of heart rate variability biofeedback in subjects with stress-related chronic neck pain: a pilot study. *Appl Psychophysiol Biofeedback*. 2011;36(2):71–80. <https://doi.org/10.1007/s10484-011-9147-0>.
  42. Whited A, Larkin KT, Whited M. Effectiveness of emWave biofeedback in improving heart rate variability reactivity to and recovery from stress. *Appl Psychophysiol Biofeedback*. 2001;4:39:75–88. <https://doi.org/10.1007/s10484-014-9243-z>.
  43. McCraty R, Atkinson M, Lipsenthal L, Arguelles L. New hope for correctional officers: an innovative program for reducing stress and health risks. *Appl Psychophysiol Biofeedback*. 2009;34:251–72. <https://doi.org/10.1007/s10484-009-9087-0>.
  44. Sutarto AP, Wahab MNA, Zin NM. Resonant breathing biofeedback training for stress reduction among manufacturing operators. *Int J Occup Saf Ergon*. 2012;18(4):549–61. <https://doi.org/10.1080/10803548.2012.11076959>.
  45. Weber CS, Thayer JF, Rudat M, Wirtz PH, Zimmermann-Viehoff F, Thomas A, et al. Low vagal tone is associated with impaired post stress recovery of cardiovascular, endocrine, and immune markers. *Eur J Appl Physiol*. 2010;109:201–11. <https://doi.org/10.1007/s00421-009-1341-x>.
  46. Kalsas GA, Chrousos GP. Cardiovascular psychophysiology. In: Cacioppo JT, Tassinary LG, Berntson GG, editors. *Handbook of psychophysiology*. 4th ed. Cambridge Handbooks in Psychology. Cambridge: Cambridge

- University Press; 2007. p. 303–18. <https://doi.org/10.1017/CBO9780511546396.013>.
47. La Marca R, Waldvogel P, Thörn H, Tripod M, Wirtz PH, Pruessner JC, et al. Association between Cold Face Test-induced vagal inhibition and cortisol response to acute stress. *Psychophysiology*. 2011;48(3):420–9. <https://doi.org/10.1111/j.1469-8986.2010.01078.x>.
  48. Mueller B, Figueroa A, Robinson-Papp J. Structural and functional connections between the autonomic nervous system, hypothalamic–pituitary–adrenal axis, and the immune system: a context and time dependent stress response network. *Neuro Sci*. 2022;1–10. <https://doi.org/10.1007/s10072-021-05810-1>.
  49. Rotenberg S, McGrath JJ. Inter-relation between autonomic and HPA axis activity in children and adolescents. *Biol Psychol*. 2016;117:16–25. <https://doi.org/10.1016/j.biopsycho.2016.01.015>.
  50. Uchino BN, Cacioppo JT, Malarkey W, Glaser R. Individual differences in cardiac sympathetic control predict endocrine and immune responses to acute psychological stress. *J Pers Soc Psychol*. 1995;69(4):736. <https://doi.org/10.1037//0022-3514.69.4.736>.
  51. 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. <https://doi.org/10.3389/fpsyg.2014.01040>.
  52. Sakakibara M, Kaneda M, Oikawa LO. Efficacy of paced breathing at the low-frequency peak on heart rate variability and baroreflex sensitivity. *Appl Psychophysiol Biofeedback*. 2020;45(1):31–7. <https://doi.org/10.1007/s10484-019-09453-z>.
  53. Weibel RP, Kerr JJ, Naegelin M, Ferrario A, Schinazi VR, La Marca R, et al. Virtual reality-supported biofeedback for stress management: beneficial effects on heart rate variability and user experience. *Comput Hum Behav*. 2023;141: 107607. <https://doi.org/10.1016/j.chb.2022.107607>.
  54. Lin G, Xiang Q, Fu X, Wang S, Wang S, Chen S, et al. Heart rate variability biofeedback decreases blood pressure in prehypertensive subjects by improving autonomic function and baroreflex. *J Altern Complement Med*. 2012;18(2):143–52. <https://doi.org/10.1089/acm.2010.0607>.
  55. Steffen PR, Austin T, DeBarros A, Brown T. The impact of resonance frequency breathing on measures of heart rate variability, blood pressure, and mood. *Front Public Health*. 2017;5:222. <https://doi.org/10.3389/fpubh.2017.00222>.
  56. Meichenbaum D. Cognitive factors in biofeedback therapy. *Biofeedback Self Regul*. 1976;1:201–16. <https://doi.org/10.1007/BF00998587>.
  57. Yu B, Funk M, Hu J, Wang Q, Feijs L. Biofeedback for everyday stress management: A systematic review. *Front ICT*. 2018;5:23. <https://doi.org/10.3389/fict.2018.00023>.
  58. Lüddecke R, Felnhöfer A. Virtual reality biofeedback in health: a scoping review. *Appl Psychophysiol Biofeedback*. 2021;47:1–15. <https://doi.org/10.1007/s10484-021-09529-9>.
  59. Andersen NJ, Schwartzman D, Martinez C, Cormier G, Drapeau M. Virtual reality interventions for the treatment of anxiety disorders: A scoping review. *J Behav Ther Exp Psychiatry*. 2023:101851. <https://doi.org/10.1016/j.jbtep.2023.101851>.
  60. Kothgassner OD, Goreis A, Bauda I, Ziegenaus A, Glenk LM, Felnhöfer A. Virtual reality biofeedback interventions for treating anxiety. *Wien Klin Wochenschr*. 2022;134(Suppl 1):49–59. <https://doi.org/10.1007/s00508-021-01991-z>.
  61. Naylor M, Ridout B, Campbell A. A scoping review identifying the need for quality research on the use of virtual reality in workplace settings for stress management. *Cyberpsychology Behav Soc Netw*. 2020;23(8):506–18. <https://doi.org/10.1089/cyber.2019.0287>.
  62. Makransky G, Petersen GB. The Cognitive Affective Model of Immersive Learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. *Educ Psychol Rev*. 2021;33(3):937–58. <https://doi.org/10.1007/s10648-020-09586-2>.
  63. Gaume A, Vialatte A, Mora-Sánchez A, Ramdani C, Vialatte FB. A psycho-engineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback. *NeuroSci Biobehav Rev*. 2016;68:891–910. <https://doi.org/10.1016/j.neubiorev.2016.06.012>.
  64. Gaggioli A, Pallavicini F, Morganti L, Serino S, Scaratti C, Briguglio M, et al. Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. *J Med Internet Res*. 2014;16(7):e167. <https://doi.org/10.2196/jmir.3235>.
  65. Blum J, Rockstroh C, Göritz AS. Heart rate variability biofeedback based on slow-paced breathing with immersive virtual reality nature scenery. *Front Psychol*. 2019;10:2172. <https://doi.org/10.3389/fpsyg.2019.02172>.
  66. Blum J, Rockstroh C, Göritz AS. Development and pilot test of a virtual reality respiratory biofeedback approach. *Appl Psychophysiol Biofeedback*. 2020;45(3):153–63. <https://doi.org/10.1007/s10484-020-09468-x>.
  67. Houzangbe S, Christmann O, Gorisse G, Richir S. Effects of voluntary heart rate control on user engagement and agency in a virtual reality game. *Virtual Reality*. 2020;24(4):665–81. <https://doi.org/10.1007/s10055-020-00429-7>.
  68. Rockstroh C, Blum J, Göritz AS. Virtual reality in the application of heart rate variability biofeedback. *Int J Hum Comput Stud*. 2019;130:209–20. <https://doi.org/10.1016/j.ijhcs.2019.06.011>.
  69. Maarsingh BM, Bos J, Tuijn CFJ, Renard SB. Changing stress mindset through stressjam: a virtual reality game using biofeedback. *Games Health J*. 2019;8(5):326–31. <https://doi.org/10.1089/g4h.2018.0145>.
  70. Kim H, Kim DJ, Kim S, Chung WH, Park KA, Kim JDC, et al. Effect of virtual reality on stress reduction and change of physiological parameters including heart rate variability in people with high stress: an open randomized crossover trial. *Front Psychiatry*. 2021;12: 614539. <https://doi.org/10.3389/fpsyg.2021.614539>.
  71. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol*. 2000;55(1):68–78. <https://doi.org/10.1037//0003-066x.55.1.68>.
  72. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. *Front Psychol*. 2018;9:1–15. <https://doi.org/10.3389/fpsyg.2018.00797>.
  73. Stenling A, Lindwall M, Hassmen P. Changes in perceived autonomy support, need satisfaction, motivation, and well-being in young elite athletes. *Sport Exerc Perform Psychol*. 2015;4(1):50–61. <https://doi.org/10.1037/spy0000027>.
  74. Shannon S, Hanna D, Leavey G, Haughey T, Neill D, Breslin G. The association between mindfulness and mental health outcomes in athletes: testing the mediating role of autonomy satisfaction as a core psychological need. *Int J Sport Exerc Psychol*. 2020:1–16. <https://doi.org/10.1080/1612197X.2020.1717578>.
  75. Alfnsson S, Johansson K, Uddling J, Hursti T. Differences in motivation and adherence to a prescribed assignment after face-to-face and online psychoeducation: an experimental study. *BMC Psychol*. 2017;5:1–13.
  76. Institute for Economics, University of Zurich. University Registration Center for Study Participants. <https://www.uast.uzh.ch/>. Accessed 04 Apr 2023
  77. Bouchery EE, Harwood HJ, Sacks JJ, Simon CJ, Brewer RD. Economic costs of excessive alcohol consumption in the U.S., 2006. *Am J Prev Med*. 2011;41(5):516–24. <https://doi.org/10.1016/j.amepre.2011.06.045>.
  78. Hirshoren N, Tzoran I, Makrienko I, Edoute Y, Plawner MM, Itskovitz-Eldor J, et al. Menstrual cycle effects on the neurohumoral and autonomic nervous systems regulating the cardiovascular system. *J Clin Endocrinol Metab*. 2002;87(4):1569–75. <https://doi.org/10.1210/jcem.87.4.8406>.
  79. Laborde S, Mosley E, Thayer JF. Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experimental planning, data analysis, and data reporting. *Front Psychol*. 2017;8. <https://doi.org/10.3389/fpsyg.2017.00213>.
  80. Mezzacappa ES, Kelsey RM, Katkin ES. Breast feeding, bottle feeding, and maternal autonomic responses to stress. *J Psychosom Res*. 2005;58:351–65. <https://doi.org/10.1016/j.jpsychores.2004.11.004>.
  81. Teare MD, Dimairo M, Shephard N, Hayman A, Whitehead A, Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *BMC Trials*. 2014;15(2):264. <https://doi.org/10.1186/1745-6215-15-264>.
  82. Twisk JWR. *Applied Multilevel Analysis: A Practical Guide for Medical Researchers*. Practical Guides to Biostatistics and Epidemiology. Cambridge University Press; 2006. <https://doi.org/10.1017/CBO9780511610806>.
  83. Kreft I, de Leeuw J. *Introducing Multilevel Modeling*. SAGE Publications, Inc.; 1998. <https://doi.org/10.4135/9781849209366>.
  84. Hox J, McNeish D. Small samples in multilevel modeling. In: van de Schoot R, Miočević M, editors. *Small sample size solutions*. 1st ed. London: Routledge; 2020. p. 215–25.
  85. 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(5):796–805.

86. Lehrer P, Vaschillo B, Zucker T, Graves J, Katsamanis M, Aviles M, et al. Protocol for heart rate variability biofeedback training. *Biofeedback*. 2013;41:98–109. <https://doi.org/10.5298/1081-5937-41.3.08>.
87. Polar Electro. SDK for Polar sensors. <https://github.com/polarofficial/polar-ble-sdk>. Accessed 27 July 2023.
88. Lalanza JF, Lorente S, Bullich R, García C, Losilla JM, Capdevila L. Methods for heart rate variability biofeedback (HRVB): A systematic review and guidelines. *Appl Psychophysiol Biofeedback*. 2023. <https://doi.org/10.1007/s10484-023-09582-6>.
89. Von Dawans B, Kirschbaum C, Heinrichs M. The Trier Social Stress Test for Groups (TSST-G): a new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology*. 2011;36:514–22. <https://doi.org/10.1016/j.psyneuen.2010.08.004>.
90. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. 1983;24(4):385–96. <https://doi.org/10.2307/2136404>.
91. Klein EM, Brähler E, Dreier M, Reinecke L, Müller KW, Schmutz G, et al. The German version of the Perceived Stress Scale - psychometric characteristics in a representative German community sample. *BMC Psychiatry*. 2016;16(1):1–10. <https://doi.org/10.1186/s12888-016-0875-9>.
92. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 2nd ed. Sydney: Psychology Foundation of Australia; 1995. <https://doi.org/10.1037/t01004-000>.
93. Nilges P, Essau C. Die Depressions-Angst-Stress-Skalen: Der DASS – ein Screeningverfahren nicht nur für Schmerzpatienten. *Schmerz*. 2015;10(4):649–57. <https://doi.org/10.1007/s00482-015-0019-z>.
94. Michalak J, Heidenreich T, Ströhle G, Nachtigall C. Die deutsche Version der Mindful Attention and Awareness Scale (MAAS) Psychometrische Befunde zu einem Achtsamkeitsfragebogen. *Z Klin Psychol Psychother*. 2008;37(3):200–8.
95. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. *Psychother Psychosom*. 2015;84(3):167–76. <https://doi.org/10.1159/000376585>.
96. WHO Regional Office for Europe: Copenhagen. Wellbeing measures in primary health care/The Depcare Project. Copenhagen: World Health Organization; 1998.
97. Morfeld M, Kirchberger I, Bullinger M. SF-36 Fragebogen zum Gesundheitszustand: Deutsche Version des Short Form-36 Health Survey. In: SF-36 Fragebogen zum Gesundheitszustand. Deutsche Version des Short Form-36 Health Survey. 2nd ed. Hogrefe; 2011. p. 1–121.
98. Smets E, Garssen B, Bonke Bd, De Haes J. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res*. 1995;39(3):315–25.
99. Heissel A, Pietrek A, Flunger B, Fydrich T, Rapp MA, Heinzel S, et al. The validation of the German Basic Psychological Need Satisfaction and Frustration Scale in the context of mental health. *Eur J Health Psychol*. 2018;25(4):119–32. <https://doi.org/10.1027/2512-8442/a000017>.
100. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*. 1996;17(3):354–81. <https://doi.org/10.1093/oxfordjournals.eurheartj.a014868>.
101. Steyer R, Schwenkmezger P, Notz P, Eid M. Der Mehrdimensionale Befindlichkeitsfragebogen MDBF [Multidimensional Mood Questionnaire]. Göttingen: Hogrefe; 1997.
102. Skoluda N, Dhrami I, Nater UM. Factors contributing to stability and instability in alpha-amylase activity in diluted saliva samples over time. *Psychoneuroendocrinology*. 2020;121:104847. <https://doi.org/10.1016/j.psyneuen.2020.104847>.
103. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. *Presence*. 1998;7(3):225–40. <https://doi.org/10.1162/105474698565686>.
104. Rheinberg F, Vollmeyer R, Engeser S. Die Erfassung des Flow-Erlebens. *Diagnostik Motiv Selbstkonzept*. 2002;2:261–79.
105. Brooke J. SUS: A 'quick and dirty' usability scale. In: Usability evaluation in industry. 1st ed. London: CRC Press; 1996. <https://doi.org/10.1201/9781498710411>.
106. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. *MIS Q*. 2003;425–78. <https://doi.org/10.2307/30036540>.
107. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. *Int J Aviat Psychol*. 1993;3(3):203–20. [https://doi.org/10.1207/s15327108ijap0303\\_3](https://doi.org/10.1207/s15327108ijap0303_3).
108. Cohen J. A power primer. *Psychol Bull*. 1992;115:5–9.
109. Field A. 14 Mixed Designs (GLM 5). In: Field A, Miles J, Field Z, editors. *Discovering statistics using R*. 1st ed. London: SAGE Publications Ltd; 2012. p. 414–5.
110. Seco GV, García MA, García MPF, Rojas PEL. Multilevel bootstrap analysis with assumptions violated. *Psicothema*. 2013;25(4):520–8.
111. Maas CJ, Hox JJ. Robustness issues in multilevel regression analysis. *Statistica Neerlandica*. 2004;58(2):127–37.
112. Hinz A, Daig I, Petrowski K, Brähler E. Mood in the German Population: Norms of the Multidimensional Mood Questionnaire MDBF. *Psychother Psychosom Med Psychol*. 2012 02;62(2):52–7. <https://doi.org/10.1055/s-0031-1297960>.
113. Kerr JI, Naegelin M, Weibel RP, Ferrario A, La Marca R, von Wangenheim F, et al. The effects of acute work stress and appraisal on psychobiological stress responses in a group office environment. *Psychoneuroendocrinology*. 2020;121. <https://doi.org/10.1016/j.psyneuen.2020.104837>.
114. González-Ramírez MT, Rodríguez-Ayán MN, Hernández RL. The Perceived Stress Scale (PSS): normative data and factor structure for a large-scale sample in Mexico. *Span J Psychol*. 2013;16:E47. <https://doi.org/10.1017/sjp.2013.35>.
115. Bech P, Olesen LR, Kjoller M, Rasmussen NK. Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 Mental Health subscale and the WHO-Five well-being scale. *Int J Methods Psychiatr Res*. 2003;12(2):85–91.
116. Avram R, Tison GH, Aschbacher K, Kuhar P, Vittinghoff E, Butzner M, et al. Real-world heart rate norms in the Health eHeart study. *NPJ Digit Med*. 2019;2(1):58.
117. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. *Front Public Health*. 2017;5:258.
118. Nunan D, Sandercock GR, Brodie DA. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *Pacing Clin Electrophysiol*. 2010;33(11):1407–17.
119. Lin JD, Chen YL, Wu CZ, Hsieh CH, Pei D, Liang YJ, et al. Identification of normal blood pressure in different age group. *Medicine*. 2016;95(14):e3188. <https://pubmed.ncbi.nlm.nih.gov/27057846/>.
120. Lüscher TF. What is a normal blood pressure? *Eur Heart J*. 2018;39(24):2233–40.
121. Heathers J. Sympathovagal balance from heart rate variability: an obituary. *Experimental Physiology*. 2012;97(4):556.
122. Gaab J, Blättler N, Menzi T, Pabst B, Stoyer S, Ehler U. Randomized controlled evaluation of the effects of cognitive-behavioral stress management on cortisol responses to acute stress in healthy subjects. *Psychoneuroendocrinology*. 2003;28(6):767–79. [https://doi.org/10.1016/S0306-4530\(02\)00069-0](https://doi.org/10.1016/S0306-4530(02)00069-0).
123. Storch M, Gaab J, Küttel Y, Stüssi AC, Fend H. Psychoneuroendocrine effects of resource-activating stress management training. *Health Psychol*. 2007;26:456–63. <https://doi.org/10.1037/0278-6133.26.4.456>.
124. Hammerfald K, Eberle C, Grau M, Kinsperger A, Zimmermann A, Ehler U, et al. Persistent effects of cognitive-behavioral stress management on cortisol responses to acute stress in healthy subjects — A randomized controlled trial. *Psychoneuroendocrinology*. 2006;31(3):333–9. <https://doi.org/10.1016/j.psyneuen.2005.08.007>.
125. de Bruin EI, van der Zwan JE, Bögels SM. A RCT comparing daily mindfulness meditations, biofeedback exercises, and daily physical exercise on attention Control, executive functioning, mindful awareness, self-compassion, and worrying in stressed young adults. *Mindfulness*. 2016;7:1182–92. <https://doi.org/10.1007/s12671-016-0561-5>.
126. van der Zwan J, Vente W, Huizink A, Bögels S, Bruin E. Physical activity, mindfulness meditation, or heart rate variability biofeedback for stress reduction: A randomized controlled trial. *Appl Psychophysiol Biofeedback*. 2015;06(40):257–68. <https://doi.org/10.1007/s10484-015-9293-x>.
127. Mather M, Thayer JF. How heart rate variability affects emotion regulation brain networks. *Curr Opin Behav Sci*. 2018;19:98–104. <https://doi.org/10.1016/j.cobeha.2017.12.017>.
128. Tops M, Buisman-Pijlman FTA, Carter CS. Oxytocin and attachment facilitate a shift from seeking novelty to recognizing and preferring familiarity: the key to increasing resilience? In: Kent M, Davis MC, Reich JW, editors. *The resilience handbook: approaches to stress and trauma*. London: Routledge Publishers; 2013. p. 115–30.
129. Wyss T, Boesch M, Roos L, Tschopp C, Frei KM, Annen H, et al. Aerobic fitness level affects cardiovascular and salivary alpha amylase responses to acute psychosocial stress. *Sports Med Open*. 2016;2. <https://doi.org/10.1186/s40798-016-0057-9>.

130. Thayer JF, Lane RD. Claude Bernard and the heart–brain connection: further elaboration of a model of neurovisceral integration. *Neurosci Biobehav Rev.* 2009;33(2):81–8. <https://doi.org/10.1016/j.neubiorev.2008.08.004>.
131. Goodman WK, Janson J, Wolf JM. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology.* 2017;80:26–35. <https://doi.org/10.1016/j.psyneuen.2017.02.030>.
132. Kloos N, Trompetter HR, Bohlmeijer ET, Westerhof GJ. Longitudinal associations of autonomy, relatedness, and competence with the well-being of nursing home residents. *Gerontologist.* 2018;59(4):635–43. <https://doi.org/10.1093/geront/gny005>.
133. Ng JYY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-Determination Theory applied to health contexts: a meta-analysis. *Perspect Psychol Sci.* 2012;7(4):325–40. <https://doi.org/10.1177/1745691612447309>.
134. Ohly S, Latour A. Work-related smartphone use and well-being in the evening. *J Pers Psychol.* 2014;13(4):174–83. <https://doi.org/10.1027/1866-5888/a000114>.
135. Fotiadis A, Abdulrahman K, Spyridou A. The mediating roles of psychological autonomy, competence and relatedness on work-life balance and well-Being. *Front Psychol.* 2019;10:1267.
136. Lubans DR, Smith JJ, Morgan PJ, Beauchamp MR, Miller A, Lonsdale C, et al. Mediators of psychological well-being in adolescent boys. *J Adolesc Health.* 2016;58(2):230–6.
137. Ryan RM, Deci EL. Chapter 2 - When rewards compete with nature: The undermining of intrinsic motivation and Self-Regulation. In: Sansone C, Harackiewicz JM, editors. *Intrinsic and Extrinsic Motivation.* Educational Psychology. San Diego: Academic Press; 2000. p. 13–54. <https://doi.org/10.1016/B978-012619070-0/50024-6>. <https://www.sciencedirect.com/science/article/pii/B9780126190700500246>.
138. Beauchamp TL, Childress JF. *Principles of biomedical ethics.* 5th ed. Oxford University Press; 2001.
139. Kerr JI, Naegelin M, Benk M, v Wangenheim F, Meins E, Viganò E, et al. Investigating employees' concerns and wishes regarding digital stress management interventions with Value Sensitive Design: mixed methods study. *J Med Internet Res.* 2023;25:e44131. <https://doi.org/10.2196/44131>.
140. Michela A, Peer JMv, Brammer JC, Nies A, Rooij MMJWv, Oostenveld R, et al. Deep-breathing biofeedback trainability in a virtual-reality action game: a single-case design study with police trainers. *Front Psychol.* 2022;13:806163. <https://doi.org/10.3389/fpsyg.2022.806163>.
141. Prabhu VG, Stanley L, Morgan R. A biofeedback enhanced adaptive virtual reality environment for managing surgical pain and anxiety. *Int J Semant Comput.* 2020;14(03):375–93. <https://doi.org/10.1142/s1793351x20400152>.
142. Rockstroh C, Blum J, Göritz AS. Combining VR and biofeedback: the effects on perceived restorativeness and presence. *J Media Psychol.* 2020;32(4):176–86. <https://doi.org/10.1027/1864-1105/a000270>.
143. Brinkmann AE, Press SA, Helmer E, Hautzinger M, Khazan I, Vagedes J. Comparing effectiveness of HRV-biofeedback and mindfulness for workplace stress reduction: a randomized controlled trial. *Appl Psychophysiol Biofeedback.* 2020;45:307–22.
144. Munafò M, Patron E, Palomba D. Improving managers' psychophysical well-being: Effectiveness of respiratory sinus arrhythmia biofeedback. *Appl Psychophysiol Biofeedback.* 2016;41:129–39.
145. Borg W. Dealing with threats to internal validity that randomization does not rule out. *Appl Psychophysiol Biofeedback.* 1984. <http://cites.eerx.ist.psu.edu/viewdoc/download?doi=10.1.1.905.2578&rep=rep1&type=pdf>. Accessed 25 July 2023.
146. Ma L, Mazidi M, Li K, Li Y, Chen S, Kirwan R, et al. Prevalence of mental health problems among children and adolescents during the COVID-19 pandemic: A systematic review and meta-analysis. *J Affect Disord.* 2021;293:78–89.
147. Weibel RP, Kerr JI, Naegelin M, Ferrario A, Schinazi VR, La Marca R, et al. Stress Management Programme with Virtual Reality-Based HRV-Biofeedback. <https://doi.org/10.17605/OSF.IO/U7CDB>. Accessed 04 Apr 2023.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

