# Cardiovascular Data Validation of the EmbracePlus PPG Wristband During VR Stress Simulation

Alp Yurdakul

Department of Psychology: University of Twente

Specialisation: Positive Clinical Psychology and Technology (PCPT)

First Supervisor: Magdalena Sikora

Second Supervisor: Thomas Vaessen

01-07-2025

Word count: 7031

# Abstract

**Background**: Psychological stress can have physiological effects on the body through its impact on cardiovascular activity (CVA), which is measurable through heart rate (HR) and heart rate variability (HRV). Researchers can utilise photoplethysmography (PPG) wearables like the Empatica EmbracePlus, which have benefits for researchers, such as usability in daily life. However, PPG devices require validation studies to be used confidently, as they are prone to data artefacts. Furthermore, the advantages brought by VR to laboratory stress research make it important to also be used in stress validation research.

**Aim**: The aim of this study is to assess the validity ,through a priori agreement levels, of the Empatica EmbracePlus PPG device in comparison to a gold standard reference device in a VR stress research setting.

**Methods:** Utilising a sample of 20 people, participants were put into a VR environment where they underwent multiple stress-related stimulus conditions whilst both devices collected CVA data from them. Utilising a predetermined framework, Bland Altman plots and line plots were created to test the validity of CVA data collected from the EmbracePlus.

**Results:** As a result, the EmbracePlus was found to be invalid in terms of HR, SD of RR interval and stressor detection when compared to a gold standard device, whilst it was found valid in terms of collecting RMSSD(HRV) data.

**Discussion:** The lack of agreement was primarily attributed to the vulnerability of the EmbracePlus to movement and speech-related issues. Researchers must be aware of said invalidities and must conduct further validation research on the EmbracePlus using sophisticated data cleaning as well as utilising an ethnically diverse large sample.

## Table of Contents

# Cardiovascular Data Validation of the EmbracePlus PPG Wristband During VR Stress Simulation

The World Health Organisation defines stress as a state of mental or psychological tension that arises from difficult situations (WHO, 2022).  Stress in day-to-day life is a common experience for many that impacts people in a multitude of different ways. A 2021 study found that 21% of the Dutch population aged 15-25 experienced frequent stress in their work lives (CBS, 2022). Another study among German residents aged 18-79 found that high levels of stress were present in 13.9% of the female population and 8.2% of the male population (Hapke et al., 2013). This adds to the weight to the fact that long-term exposure to stress can have negative effects on people, increasing the chances of various somatic conditions, such as irritable bowel syndrome, tension headaches, and coronary heart disease (Nakao, 2017; Glise et al., 2014; Steptoe & Kivimäki, 2012).

Psychological stress affects systems in the body, such as the cardiovascular system, causing noticeable changes in autonomous nervous system (ANS) regulation, which can cause a noticeable effect on cardiovascular activity (CVA), such as in heart rate (HR) inter-beat-intervals (IBI) and heart rate variability (HRV) (O'Connor et al., 2020; Mühlen et al., 2021; Van Lier et al., 2019). An example of an explanation of these affects is due to the body reacting to stress by activating the sympathetic-adrenal-medullary system (SAM), triggering activation of the sympathetic nervous system (which elevates physiological response) whilst suppressing the parasympathetic nervous system (which cools physiological response), releasing noradrenaline as a threat reaction measure (O'Connor et al., 2020; Dunlavey, 2018; Kim et al., 2018).

## The Need for Validation Studies

Much interest has been given to measuring stress effects with CVA. Multiple studies have engaged with measuring the physiological effects of stress to both attempt to measure how people perceive stress as well as measure how it affects aspects of the human body, such as the cardiovascular system (Ahmed et al., 2023; Barnett et al., 1997; Beh, 1998; Li, 2024). In such studies, different laboratory-based measurement devices such as the electrocardiogram (ECG) are used to collect stress-related CVA data in a laboratory setting, such as HR and HRV (through methods like RMSSD); however, these devices can take time to set up and are prone to creating artificial experimental procedures, which hampers the validity of studies (Menghini et al., 2019; Mühlen et al., 2021). To counteract this, wearable devices such as the Empatica EmbracePlus wristband can be used to measure CVA using methods such as Photoplethysmography (PPG) that enables the collection of data outside of the laboratory, allowing for higher ecological validity as the data can be collected in real-life scenarios (Menghini et al., 2019; Mühlen et al., 2021).

However, as these devices tend to lack transparency in their effectiveness, their validity in comparison to gold standard ECG laboratory methods can at times be poor and is unknown without third-party research backing. (Menghini et al., 2019; Mühlen et al., 2021). For example, in a study conducted by Gerboni et al. (2023), the Empatica EmbracePlus showed high clinical validity while collecting oximeter data specifically in conditions with no motion when compared to a gold standard device. Furthermore, a study by Coelli et al. (2024) validating the Embrace within CVA parameters also found that it was significantly affected by arm movement.

Additionally, the speed at which these wearables are updated and introduced into the market creates high demand for constant validation studies. Moreover, PPG sensors can be unreliable in conditions where the participant wearing the sensor moves their wrist around, which can cause motion-related artefacts in data, making the interpretation of the BVP signal (the signal collected by PPG to measure CVA) difficult. Finally, participants having dark skin or large wrists can also potentially affect the collected data due to differences in light absorption, which the PPG depends on to function. Due to these reasons, validating wearable PPG devices in a stress measurement setting is crucial to be able to conduct stress-related research (Kleckner et al., 2020; Menghini et al., 2019; Mühlen et al., 2021).

## Stress Validation Research and VR

Previous attempts at validation of PPG wearables have been made, such as multiple 2019 studies by Menghini et al. (2019) and Van Lier et al. (2019). However, these studies did not validate the EmbracePlus and were conducted in a non-virtual reality (VR) laboratory environment.

VR is a computer-generated 3D simulation of a real-world environment, allowing participants to enter and interact with it using specialised goggles and controllers (Damgrave, 2016). Research indicates that incorporating VR environments into laboratory stress studies enhances ecological validity. According to Parsons (2015), utilising VR within affective research enables an extra level of immersion while being able to preserve laboratory environments. Analysis by Finseth et al. (2024) on the usage of VR in specifically stress-related research concurs that it allows for an increase in ecological validity in a laboratory environment through providing the participant with environment-related affordances. An additional study found that immersive VR environments could potentially be used in studies to classify types of experienced stress (Cho et al., 2017). Finally, VR also allows for the usage of more varied stimuli, such as "walking the plank", where participants walk on a virtual plank high in the air to induce stress and emotional response (Basbasse et al., 2023; Martens et al., 2019). All in all, VR has the potential to be useful to enhance ecological validity in different types of stress-related research but has mostly lacked usage in stress-related validation studies.

## Current Study

Due to the importance of validating wearable PPG devices so that they can be used for future stress related research and the benefits of using VR in stress related research, this study aims to uncover the extent to which CVA data collected from the EmbracePlus is valid when compared to data collected from the gold standard ECG reference device, in a virtual reality stress experiment setting.

Due to previous findings, it can be expected that the device will show disagreement when compared to a gold standard device during VR stress experiment conditions, where participants will not have their arm movements limited (Coelli et al. 2024). Referencing the framework of Van Lier et al. (2019), which will be explained in the methods section, the following hypotheses are made:

H1: Average PPG, HR, and standard deviation of RR interval data collected from the Empatica Embraceplus will show less than satisfactory agreement according to a priori boundaries when compared to data collected from the gold standard ECG device when tested throughout the duration of the experiment.

H2: PPG RMSSD data collected from the Empatica Embraceplus will show less than satisfactory agreement according to a priori boundaries when compared to data collected from the gold standard ECG device when tested throughout the duration of the experiment.

H3: PPG HR data collected from the Empatica Embraceplus will show less than satisfactory agreement according to a priori boundaries when compared to data collected from the gold standard ECG device, when testing for the detection of stressors, and when compared to baseline measurements.

# Methods

## Design

Data to test the hypotheses was collected through a larger laboratory study aiming to validate wearable biometric devices in comparison to gold standard research devices. Remaining within the context of this thesis and hence excluding irrelevant information, the study used a within-participant design to simultaneously collect gold standard ECG and wearable PPG(EmbracePlus) data from participants whilst being put through stress-related conditions in a VR environment. Following this, the data was analysed and compared using the framework of Van Lier et al. (2019), described below, to deduce the accuracy of the PPG data and answer the hypotheses.

## Framework

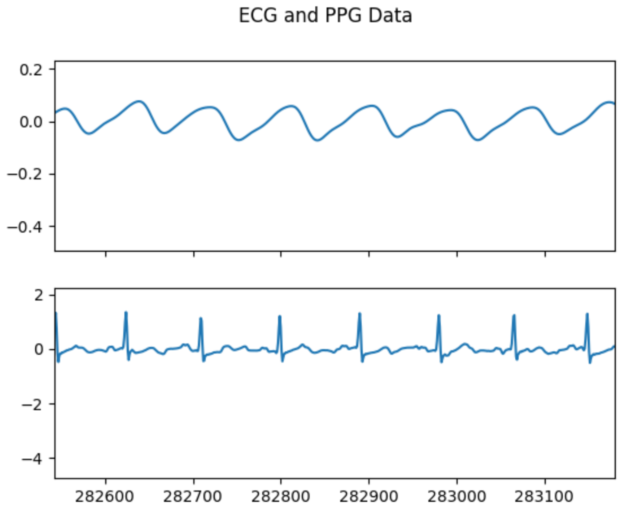
Van Lier et al. (2019)'s framework for validation was used in this study. According to Van Lier et al. (2019) 's guidelines for validating wearable devices, a wearable device can be validated by comparing data collected by it to data collected by a gold standard reference device that measures the same parameter at the same time and in the same environment. This validation can be done in 3 levels, namely the signal level, event level and parameter level.

### *Signal Level*

At the signal level, the raw data collected by the telemetric device and the reference device are compared with the degree of similarity of the values being considered (Van Lier et al., 2019). A signal level comparison is not possible for this study as the ECG and PPG devices use different methods of collecting data. This is due to QRS waves being incomparable to BVP waves and vice versa (Van Lier et al., 2019). This can be seen in Figure 1.

**Figure 1**

Comparison of QRS and BVP Signal



*Note.* Image demonstrating the difference in raw data for the PPG BVP signal(top) and the ECG QRS signal(bottom)

***Parameter Level***

At the parameter level, aggregated data per participant from the telemetric device and reference device are compared. Within the scope of CVA, HR, standard deviation of the RR/PP intervals, which are the gaps between the peaks of the waves as seen in Figure 3, and time domain parameters such as RMSSD can be compared using bland Altman plots to look for missing agreement and see if systematic biases can be seen between the devices (Van Lier et al., 2019).

### *Event Level*

In the event level, data gathered from both devices specific to the different phases of the experiment, such as when different stress stimuli and baselines occur, are compared. Using graphical visualisations, researchers interpret the extent to which the telemetric device detected changes in CVA that were in line with the reference device (Van Lier et al., 2019).

## Participants

Participants were recruited mainly through convenience sampling and were generally either friends or acquaintances of the researchers. A sample size of 20 was used. It was ensured through self-report that participants were not diagnosed with a heart disease and were not currently using medication that would influence the autonomic nervous system, such as Selective Serotonin Reuptake Inhibitors (SSRIs). All participants were exposed to the same conditions and wore the same wearables and reference devices. Participants were also instructed to sleep for 6 hours minimum and refrain from consuming alcohol, caffeine, or smoking before the study.

## Materials

### *Wearable PPG*

Empatica Embrace plus smart watch with a PPG sampling rate of 64hz was used to collect PPG data (Empatica, 2025). Additionally, a smartphone was used to utilise the watch's Empatica Care Lab app to monitor the data, which was then downloaded from the app using Cyberduck software (GmbH, 2025; Empatica, 2025b).

### *Reference device*

A Biopac MP160 ECG smart amplifier device with a sampling rate of 2000 Hz was used as the gold standard reference device (RD) (*ECG Electrocardiogram Smart Amplifier*, 2025). Skin prep gel and electrode gel were also utilised to prep the participant's skin before electrode placement and to improve the effectiveness of the electrodes (*Electrode Gel, 250 G*, 2024;

*EEG/ECG/EMG/EOG Prep Gel 114 G*, 2024). Electrodes were placed in a modified Lead II position.

### *Monk Skin Tone Scale*

The Monk skin tone scale poster was used to compare the skin on the palm side of the participant's arm to the tones on the poster to quantify participant skin colour (Monk, 2023).  The Monk skin tone scale rates skin colour from 1-10, with one being the lightest and 10 being the darkest skin tone (Monk, 2023).

## Virtual Reality Environment (Conditions)

The Meta Quest 3 VR headset and controller were worn by the participant during the VR conditions. The virtual reality environment and conditions were made and controlled with the Unity game engine (*VR Game Development Software & Engine | UNITY*, n.d.). While in the VR environment, the participant was exposed to 3 stress-related conditions. These are the TSST, walking the plank and natural environmental conditions. The TSST and plank conditions were stressors with order randomised per participant, whereas the nature condition was used for recovery and movement elicitation.

### *TSST*

The Trier Social Stress Test (TSST) is a stressor used to trigger acute stress responses in the participant. It involves the participant conducting a self-made interview presentation after a preparation phase to a panel of judges who at no point encourage the participant (Kirschbaum et al., 1993). After the presentation, the participant is made to perform an unexpected arithmetic mathematical task whilst still in front of the same panel (Allen et al., 2016; Kirschbaum et al., 1993). In the study, the environment for the test was fully in VR with virtual judges in a virtual room. All phases lasted 5 minutes, with the virtual judge verbally reminding the participant that they had more time if they were silent for too long during the presentation. The mathematical task was counting backwards from the number 1022 in intervals of 7. If a mistake was made or the participant was counting slowly, the judge verbally warned them of this and asked them to start the task over. The verbal interaction of the judges with the participant was controlled by researchers monitoring the study from a nearby computer.

### *Walking the plank*

A plank condition was used to trigger stress responses like those used in previous research to trigger stress (Basbasse et al., 2023; Martens et al., 2019). The participant was instructed to step into an elevator going up in the VR environment. Once the participant reached the top, the elevator door opened, revealing a wooden plank hanging over a ledge of a skyscraper surrounded by other skyscrapers in an urban setting. The participant was instructed by the environment via recording of verbal instructions to step onto the plank and look down. Once 2 minutes had elapsed, the participant was instructed to step back into the elevator before being taken back down.

### *Nature environment*

In the nature environment condition, the participant was put into a VR forest and instructed by the environment to walk around and collect butterflies flying around and landing on the ground. This was done, and the participants were able to touch the butterflies using their VR controller to complete the task. After 5 minutes had elapsed, the participant was automatically removed from the environment. The task aimed to both calm the participants after the stressor tasks but also introduce movement through the butterfly collection.

## Procedure

For each individual session, participants were greeted upon arrival and briefed on the nature of the study, without going into detail about the study conditions. They were then asked if they had any questions regarding the informed consent and preliminary questionnaire forms, they had previously completed. Following this, physiological data such as height, weight and skin colour were gathered, and some biometric wearable devices were attached to the participant and set to record, including the EmbracePlus and other devices for use in other studies. Next, the participant was directed to go up and down a small flight of stairs 5 times to increase heart rate and standardise baseline measurements. After this, the gold standard reference devices, including the ECG RD, were attached to the participant utilising skin gels. Data was then synchronised, and all devices were set to record. Next, baseline data was collected through having the participant stand, sit, lie down and then perform a read-aloud task for 2 minutes each. After this, the participant was put into a VR environment and given a controller before being exposed to the *TSST* condition and a *walking the plank* condition in random order. Instructions were provided for the participants within the VR environment with audio and text. Following the two conditions, the participant was placed in the third *natural environment* condition. 2-minute pauses were taken in between the conditions, with short self-report stress scales also being presented within the VR environment, which were answered by participants via a VR controller. After completion, the participant was pulled out of the VR environment and was instructed to fill out further questionnaires. Finally, participants were verbally debriefed, followed by the removal of all measurement devices and the saving of recorded data. Participants were also given the opportunity to review their collected data if they wished to do so.

## Data Analysis

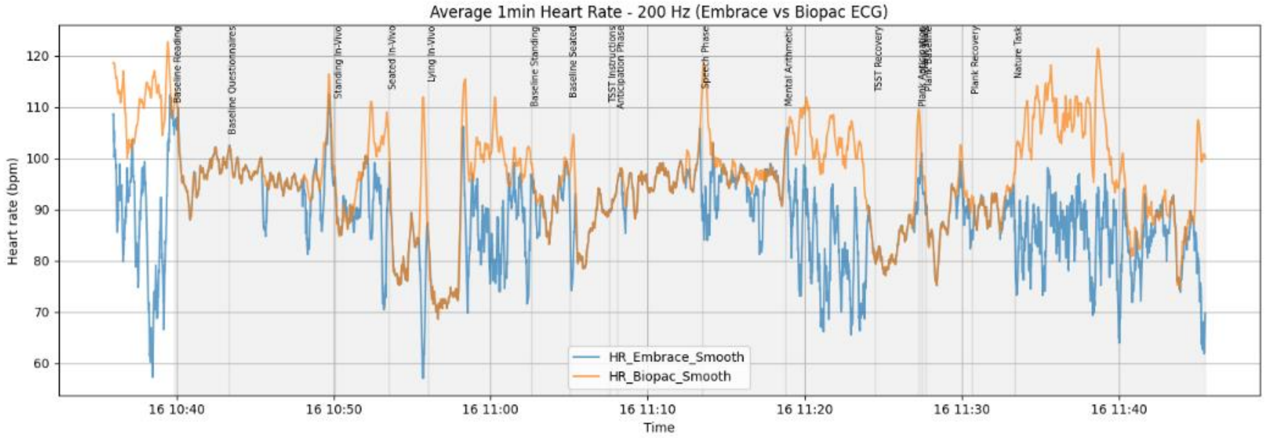
Analysis of the data was conducted utilising the framework outlined by Van Lier et al. (2019) for conducting parameter and event-level validation comparisons between ECG and PPG data.

### *Preprocessing data*

Using the Neurokit2 library in Python, the data from the BIOPAC RD and the EmbracePlus PPG were resampled to a common frequency of 200hz. The data from both devices was converted into a csv. Files per participant with RR intervals, HR values (Including average values per minute of data) and instantaneous HR values within a 15-second rolling window per device. For ECG RR intervals, the neurokit2 package in Python was used using a method based on frequency band detection for QRS waves from a study by Elgendi et al. (2010) (Makowski et al., 2021). RR intervals for the PPG data were acquired using the Empatica Care Lab software (Empatica, 2025b). Local event markers, such as "TSST", which mark the stage of the experiment where this data was collected, were also included in the file, along with event phase markers like "TSST Instructions", indicating the phase of the stage the experiment was in. Additionally, all data collected before the beginning of the first experimental condition were excluded because they fell outside the experimental conditions and were considered irrelevant (see Figure 2).

**Figure 2**

*Pre-processed PPG and ECG Data Example*



*Note*. Example of CVA data collected for 1 participant in the study. The event markers for stages of the experiment are marked above as "Baseline Questionnaire", "Baseline Reading", etc. Data to the left of the “Baseline Questionnaire” or “16 10:40” time marker can be seen as falling outside of the experimental conditions.

### *Parameter Level Analysis*

Bland-Altman plots determining the difference between the Embrace Plus and RD devices were plotted separately for the parameters of Mean HR, Mean RR intervals and RMSSD for heart rate variability.  R studio with “ggplot2” and “dplyr” was utilized to process the data and create the plots. All plots were visually analysed to determine parameter level validity with a tolerance range of ±10% data points lying within plausible values (which differ by parameter) being considered valid as per the a priori suggestion of Van Lier et al. (2019).

#### **HR**

The HR parameter was plotted by calculating the average heart rate per participant per device before calculating the difference in the averages per participant. Embrace-RD differences were then plotted against their means by means of a Bland-Altman plot. The data normality of the differences between the devices was also checked through a histogram. The Bland-Altman plot was created with 95% limits of agreement (mean bias ± 1.96 × sd of differences), mean bias, as well as boundaries for plausible HR differences, which were set at +-5bpm as per the suggestion of Van Lier et al. (2019).

#### **SD of RR Intervals**

The SD of RR intervals parameter was plotted by calculating the average SD of RR intervals per participant per device before calculating the difference of the averages per participant. Embrace-RD differences were then plotted against their means through a Bland-Altman plot. The data normality of the differences between the devices was also checked through a histogram. The Bland-Altman plot was created with 95% limits of agreement (mean bias ± 1.96

× sd of differences), mean bias as well as boundaries for plausible SD of RR interval differences, which were set at +-0.06 seconds as per the suggestion of Van Lier et al. (2019).

### RMSSD

RMSSD values per device per participant were first calculated with their respective RR values in R Studio using the RMSSD formula:

$$RMSSD = \sqrt{[(1 / (N - 1)) \times \Sigma(RR_{i+1} - RR_{i})^{2}]}$$

Using the RMSSD value per device per participant, a Bland-Altman plot was created using the same methods as in the HR and RR mentioned above. Plausible RMSSD interval differences, which were set at +-0.07 seconds as per the suggestion of Van Lier et al. (2019).

### *Event Level Analysis*

An event-level analysis was also carried out using instantaneous HR (per 15-second window) data per participant per device as suggested by Van Lier et al. (2019). Like the parameter level analysis, this data was also checked for normality between the differences of the RD and the PPG. Average instantaneous HR data per device per participant were compared per stage of the TSST and the Plank phases, with a seated and standing baseline of the experiment through multiple line graphs. These stages (event markers) included a VR seated baseline, TSST Instructions, Anticipation Phase, Speech Phase, Mental Arithmetic and recovery stage for the TSST. The marked stages for the plank task were Baseline Standing, Plank Anticipation, Plank Walk, Plank Baseline, and a recovery stage.

Using R Studio with “tidyverse” packages, an event difference line plot showcasing average instantaneous HR on the y axis and the event markers on the x axis was plotted for the Embrace Plus and RD, respectively. On these plots, each participant was given a line along with a line for the overall mean and SE per task of all participants. The purpose of these plots was to determine if the separate devices detected a change between the seated baseline and the TSST at each stage and recovery. The graphs were evaluated based on whether the SE bars per TSST phase overlapped with the VR seated baseline phase. Following the advice of Van Lier et al. (2019), if said overlapping was detected, then it was deduced that the stressor event was not detected.

Next, another line plot was created with average instantaneous HR differences (Embrace-RD) on the y axis and event markers on the x axis. According to Van Lier et al. (2019), this plot is only required if both devices detect a change in effect between the baseline and the stressor. However, even in this situation, the plot was created anyway for illustrative purposes.

On this plot, each participant was given a line along with a line for the overall mean and the SE per task of all participants. Additionally, an a priori boundary is a set of maximum valid differences of the average instantaneous HR between the Embrace Plus and reference device. This a priori limit is determined by the effect of the stress condition compared to the baseline for the RD as per the suggestion of Van Lier et al. (2019). The a priori boundaries were not included if the RD did not detect stress in any of the event markers.

## Results

### Participant Results

Of the complete sample size of 20 people, two were disqualified due to medical reasons and data usability issues. The sample size used was 18 people, 38.9% males (N=7) and 62.1% (N=11) females. The median age of the sample was 24 (lower quartile = 22, upper quartile = 25). Additionally, the average skin tone was three on the Monk skin tone scale (lower quartile = 3, upper quartile = 3), indicating mostly homogenous skin tone among participants.

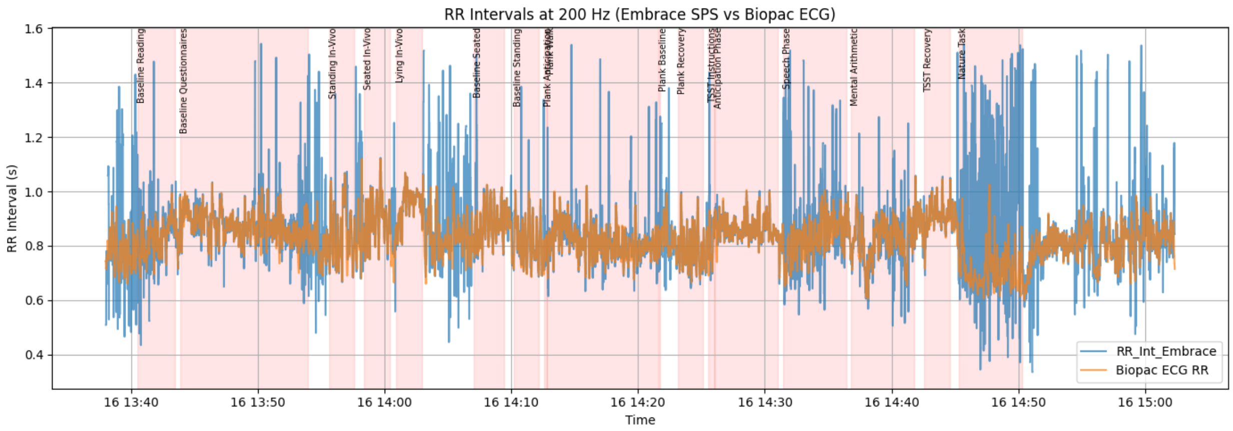
### Parameter Level Analysis

For average HR, the plot shows poor agreement between the EmbracePlus and the RD, with 44.44% of values falling outside of the acceptable +-5bpm difference boundary (see Figure 4). However, agreement is visible in average HR values lower than 80, suggesting validity within this range. For the SD of RR intervals, the plot shows poor agreement between the EmbracePlus and the RD, with 55.56% of values falling outside of the acceptable +-0.06 second difference boundary (see Figure 5). Also, large amounts of noise were detected in the RR interval data for both devices, suggesting the quality of the data may affect the accuracy of the validity statements (see Figure 3).

The first hypothesis regarding average PPG HR and SD of RR intervals data collected from the Embraceplus not being in satisfactory agreement when compared to data collected from the gold standard ECG device, when tested for the parameter level, was accepted.

**Figure 3**

*RR Intervals Embrace and RD*

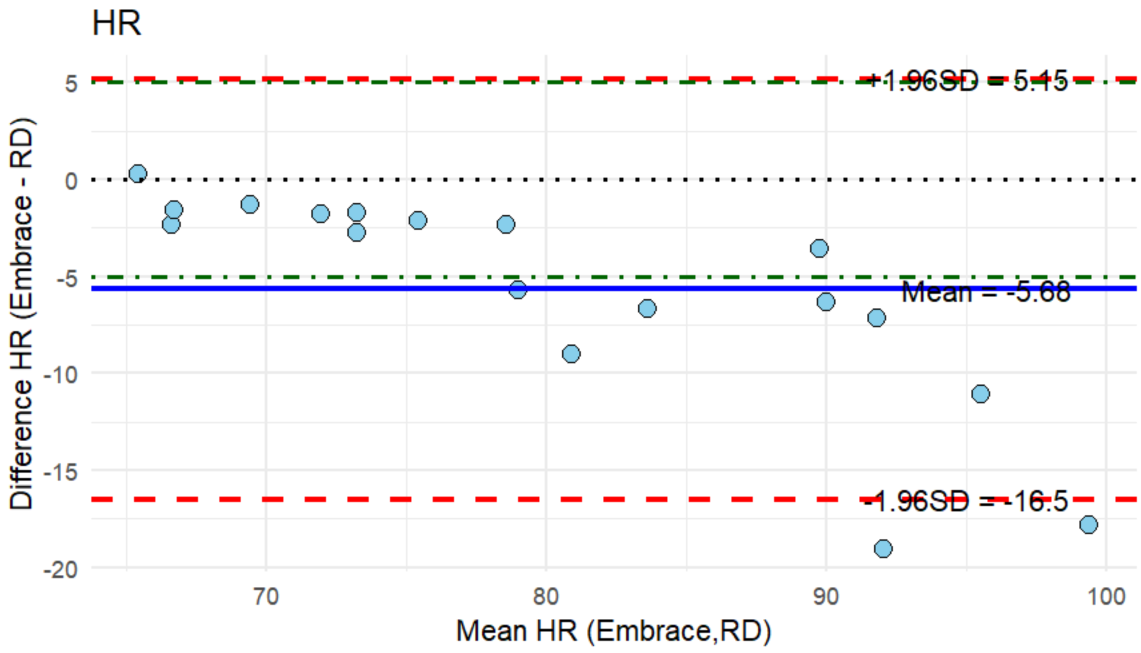


*Note.* RR intervals throughout the experimental conditions are seen for the RD(orange) and the EmbracePlus(blue). RR intervals in seconds are plotted on the y-axis while the experimental timestamp is plotted on the x-axis. Large amounts of noise can be observed in the Embrace Plus data, particularly in conditions with high participant activity.

For RMSSD, the plot shows good agreement between EmbracePlus and the RD, with 100% of values falling within the acceptable ±0.07 seconds difference boundary (see Figure 6). The RMSSD values were (M= 4.78ms, SD= 2.67ms) for the RD and (M=21.24ms, SD=7.52ms) for the EmbracePlus. As a result, the second hypothesis regarding PPG RMSSD data collected from the Embraceplus not being in satisfactory agreement when compared to data collected from the gold standard ECG device when tested for the parameter level is rejected.

**Figure 4**

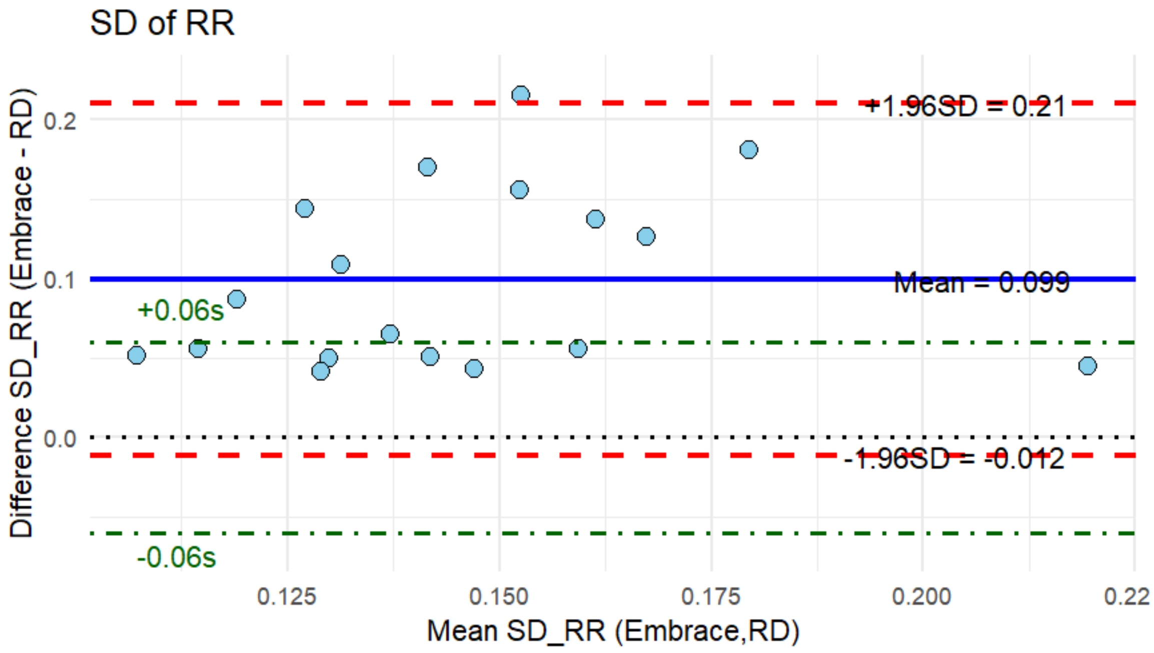
*Bland-Altman Plot Average HR*



*Note.* Every participant is represented by a dot with a difference between the EmbracePlus and the RD on the x axis, with averages on the y axis. The red dotted lines represent % 95% CI limits. The green dotted lines represent a priori boundaries of satisfactory validity. The blue line represents the mean difference or bias. At 80bpm, the EmbracePlus begins to significantly underestimate ECG HR readings.

**Figure 5**

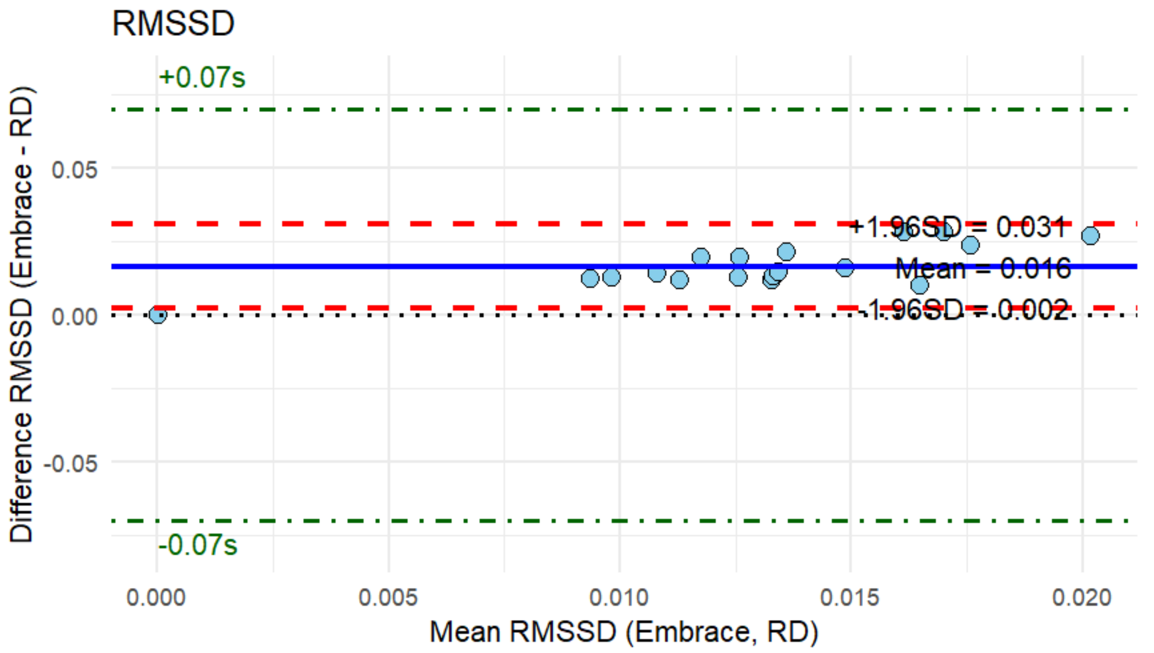
*Bland-Altman Plot SD of RR Intervals*



*Note.* See note on Figure 4

**Figure 6**

*Bland-Altman Plot RMSSD*



*Note.* See note on Figure 4

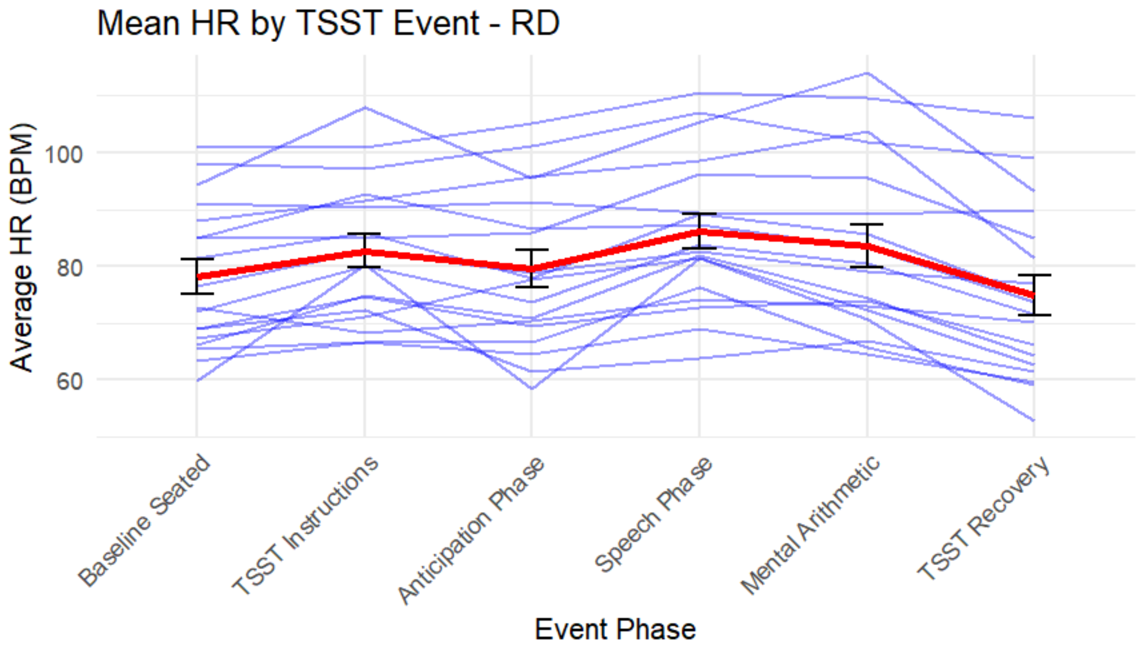
## Event Level Analysis

### *TSST Results*

RD SE bar overlapping was calculated per event phase, with the only event phase not overlapping with the baseline (95% CI Range = 75.08 – 81.12) being found as the Speech Phase (95% CI Range = 82.97 – 89.23) (see Table 1 and Figure 7). As a result, the RD detected the stressor in the Speech Phase but not in the other phases. EmbracePlus baseline overlapping was calculated per event phase, with all event phases overlapping with the baseline (95% CI Range = 74.19 – 80.01) (see Table 2 and Figure 8). Further disagreement was observed in Figure 9, with a mean HR difference for the Speech Phase of -9.13, which fell outside the ±8 bpm boundary; however, the rest of the measurements fell within the boundaries.

**Figure 7**

*Event Level TSST Line Plot RD*



*Note.* Each participant is represented with a blue line. The mean and SEs are represented with red and black lines, respectively.

**Table 1**

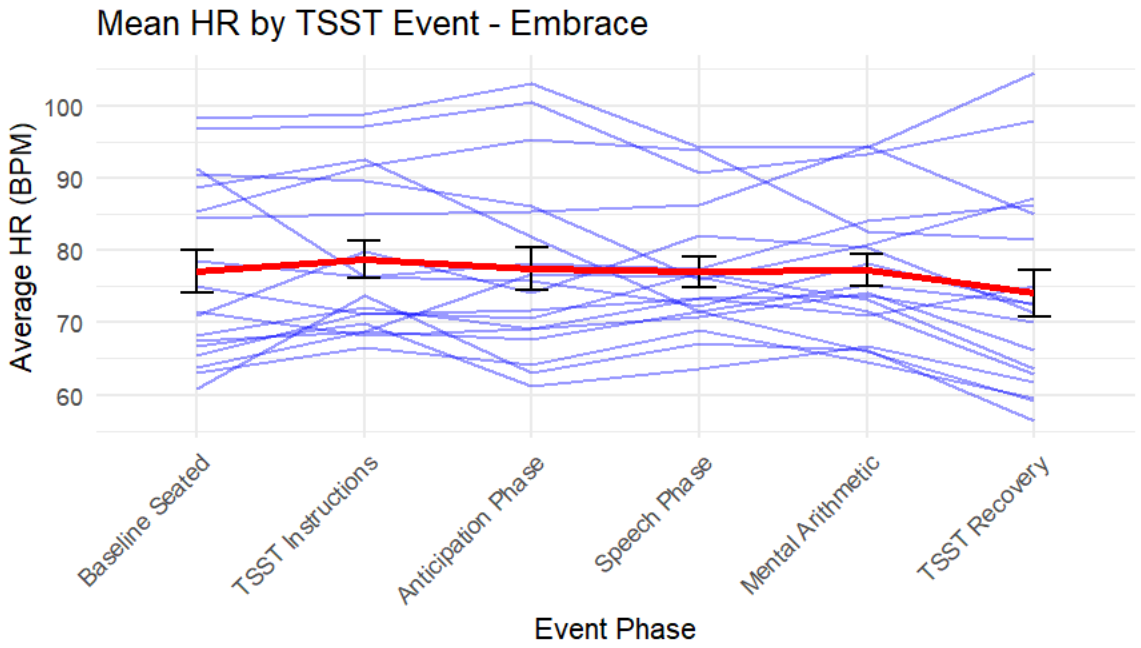
*RD Overlapping of Baseline with TSST Phases*

| Event | Mean HR (bpm) | SE | 95% CI Range | Baseline Overlapping | Stressor Event Detection |
|---|---|---|---|---|---|
| Baseline Seated | 78.1 | 3.02 | 75.08 – 81.12 | — | — |
| TISST Instructions | 82.7 | 2.88 | 79.82 – 85.58 | Yes | No |
| Anticipation Phase | 79.5 | 3.31 | 76.19 – 82.81 | Yes | No |
| Speech Phase | 86.1 | 3.13 | 82.97 – 89.23 | No | Yes |
| Mental Arithmetic | 83.5 | 3.66 | 79.84 – 87.16 | Yes | No |
| TSST Recovery | 74.8 | 3.52 | 71.28 – 78.32 | Yes | No |

*Note.* The overlapping of SE bars for the separate TSST event phases with the Baseline Seated condition was checked using the "Stressor Event Detection" column, which indicated whether the RD was able to detect a difference in Instantaneous HR compared to the Baseline Seated condition.

**Figure 8**

*Event Level TSST Line Plot EmbracePlus*



*Note.* Each participant is represented with a blue line. The mean and SEs are represented with red and black lines, respectively.

**Table 2**

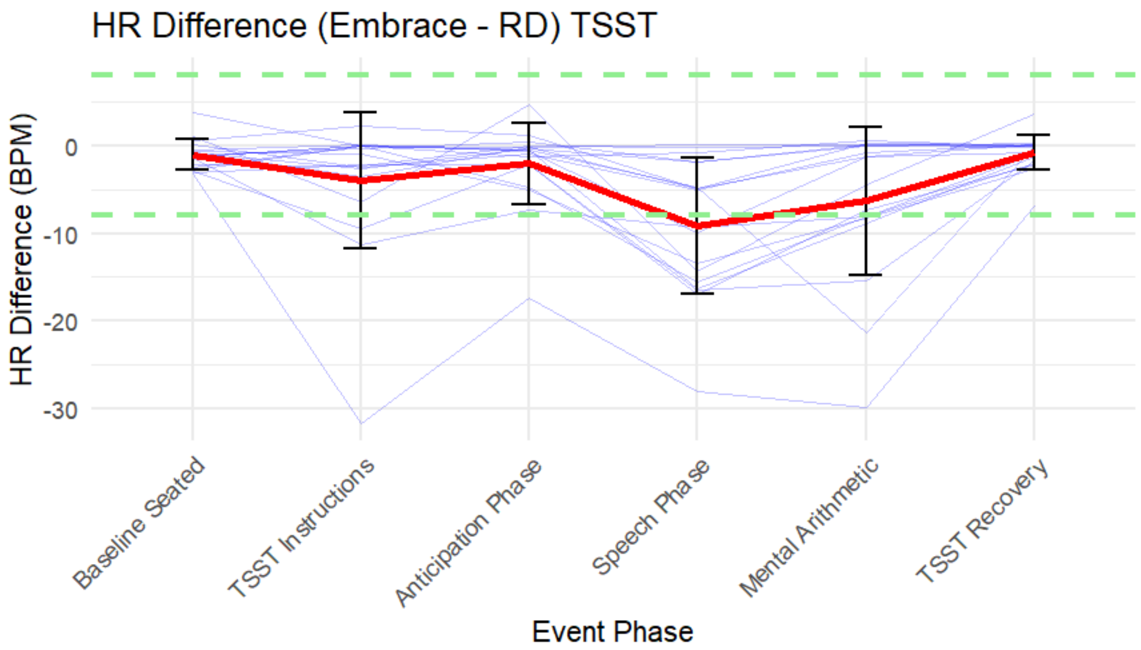
*EmbracePlus PPG Overlapping of Baseline with TSST Phases*

| Event | Mean HR (bpm) | SE | 95% CI Range | Baseline Overlapping | Stressor Event Detection |
|---|---|---|---|---|---|
| Baseline Seated | 77.1 | 2.91 | 74.19 – 80.01 | — | — |
| TISST Instructions | 78.8 | 2.55 | 76.25 – 81.35 | Yes | No |
| Anticipation Phase | 77.5 | 2.94 | 74.56 – 80.44 | Yes | No |
| Speech Phase | 77.0 | 2.12 | 74.88 – 79.12 | Yes | No |
| Mental Arithmetic | 77.2 | 2.26 | 74.94 – 79.46 | Yes | No |
| TSST Recovery | 74.1 | 3.24 | 70.86 – 77.34 | Yes | No |

*Note.* The overlapping of SE bars for the separate TSST event phases with the Baseline Seated condition was checked using the “Stressor Event Detection” column, which indicated whether EmbracePlus was able to detect a difference in Instantaneous HR compared to the Baseline Seated condition.

**Figure 9**

*Event Difference Line Plot TSST*



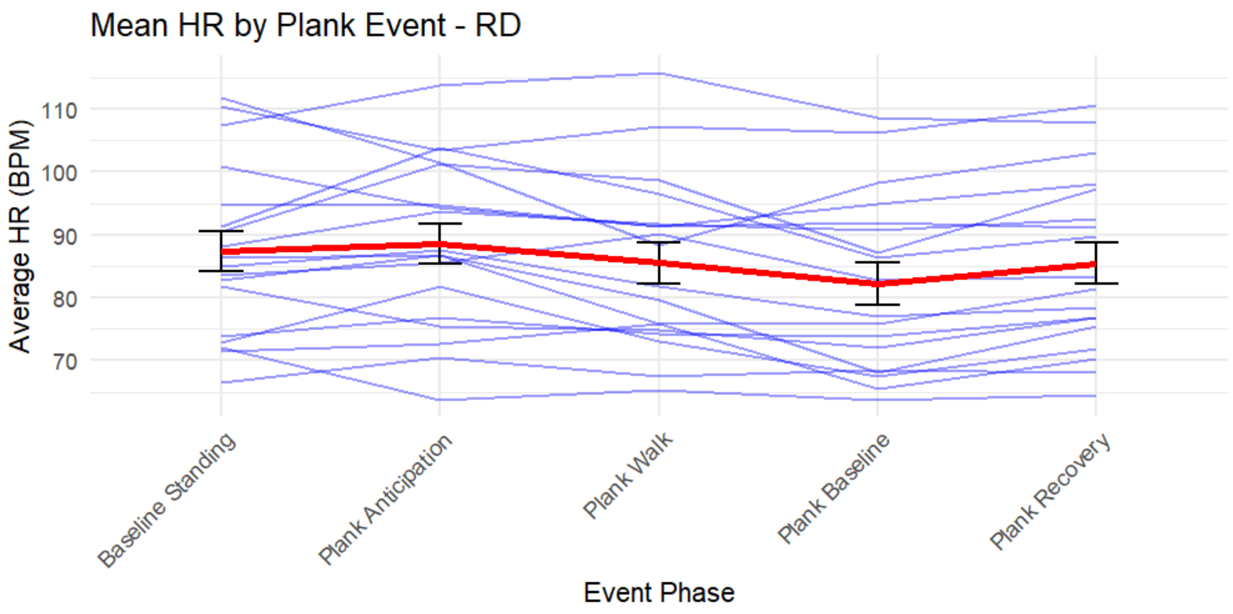
*Note.* Instantaneous HR difference (Embrace-RD) is shown by event. Each participant is represented with a blue line. The mean and SEs are represented with red and black lines, respectively. A priori boundaries of +-8bpm are represented by green dotted lines. The lack of agreement is evident during the "Speech Phase," where the EmbracePlus fails to detect the stressor, resulting in underreporting of HR compared to the RD.

### ***Plank Results***

RD SE bar overlapping was calculated per event phase, with overlapping being found between the baseline (95% CI Range = 84.07 – 90.53) and the stressors for all phases of the plank condition (see Table 3 and Figure 10). As a result, the RD did not detect the stressor. The Embrace Plus detected a change in instantaneous HR between the baseline (95% CI Range = 81.38 – 86.22) and all phases outside of the recovery phase; however, these findings are ignored as the RD did not detect any of these stressors. Additionally, difference between the devices is seen with the mean difference of multiple events exceeding -8 (the a priori boundary for the TSST) such as Plank Anticipation (M = -11.3, SD =11.1 ) and Plank Walk (M = -9.05, SD = 10.8) (see Figure 9 and 13).

**Figure 10**

*Event Level Plank Line Plot RD*



*Note.* Each participant is represented with a blue line. The mean and SEs are represented with red and black lines, respectively.

**Table 3**

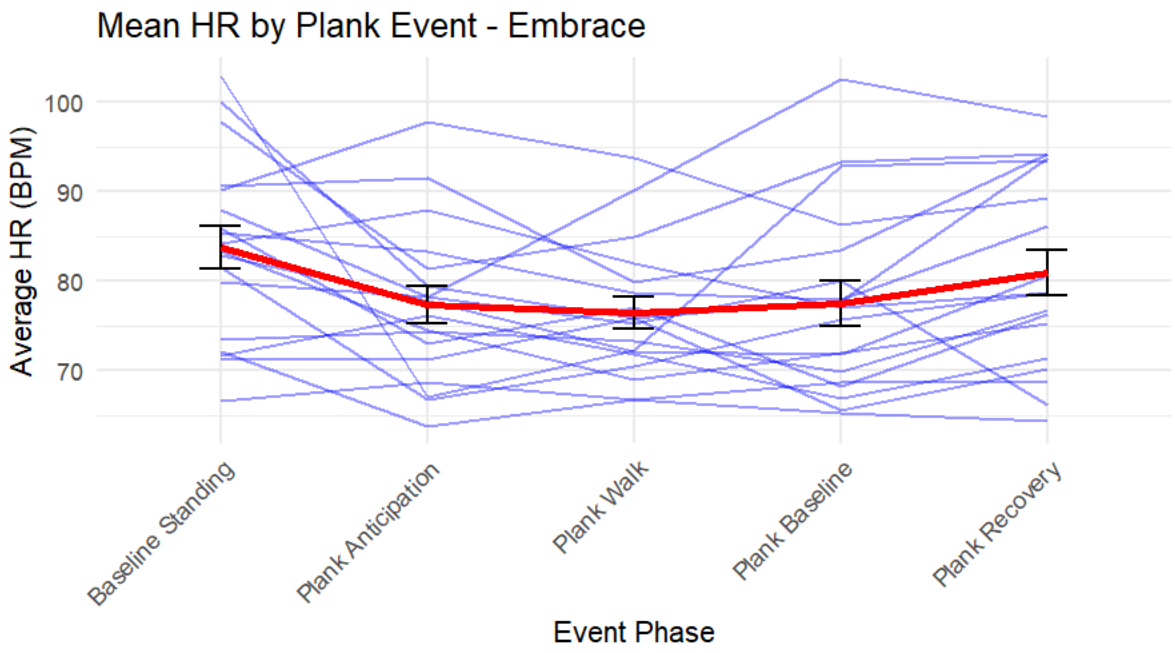
*RD Overlapping of Baseline with Plank Phases*

| Event | Mean HR (bpm) | SE | 95% CI Range | Baseline Overlapping | Stressor Event Detection |
|---|---|---|---|---|---|
| Baseline Standing | 87.3 | 3.23 | 84.07 – 90.53 | — | — |
| Plank Anticipation | 88.6 | 3.18 | 85.42 – 91.78 | Yes | No |
| Plank Walk | 85.5 | 3.23 | 82.27 – 88.73 | Yes | No |
| Plank Baseline | 82.2 | 3.32 | 78.88 – 85.52 | Yes | No |
| Plank Recovery | 85.4 | 3.30 | 82.10 – 88.70 | Yes | No |

*Note.* Overlapping of SE bars of the separate TSST event phases with the Baseline Standing condition was checked with the "Stressor Event Detection" column indicating whether the RD was able to detect a difference in Instantaneous HR when compared to the Baseline standing condition.

**Figure 11**

*Event Level Plank Line Plot EmbracePlus*



*Note.* Each participant is represented with a blue line. The mean and SEs are represented with red and black lines, respectively.

**Table 4**

*EmbracePlus Overlapping of Baseline with Plank Phases*

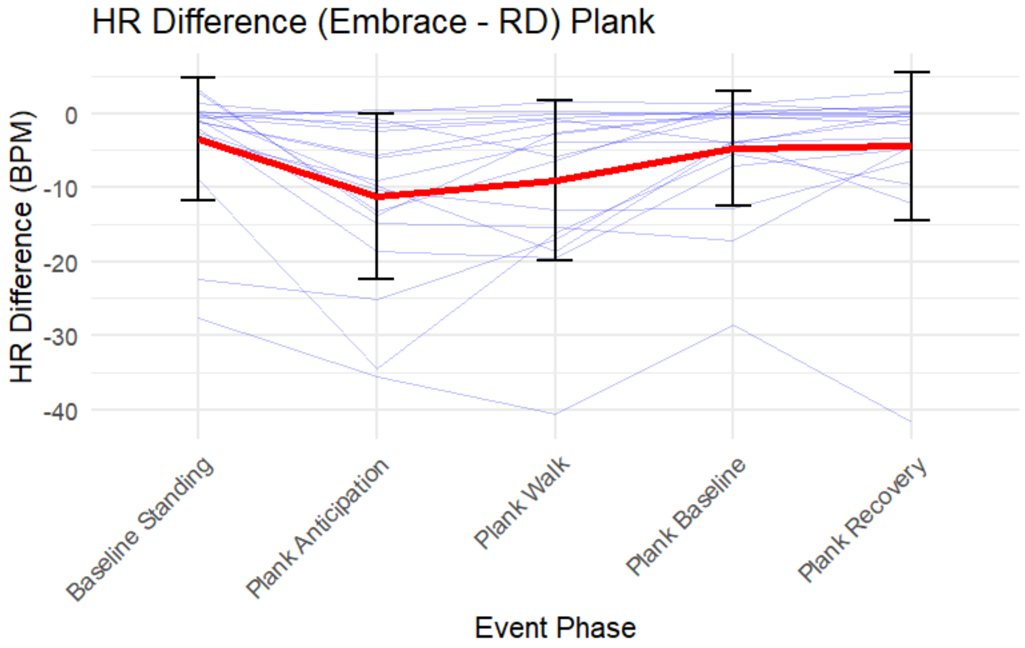
| Event | Mean HR (bpm) | SE | 95% CI Range | Baseline Overlapping | Stressor Event Detection |
|---|---|---|---|---|---|
| Baseline Standing | 83.8 | 2.42 | 81.38 – 86.22 | — | — |
| Plank Anticipation | 77.3 | 2.09 | 75.21 – 79.39 | No | Yes |
| Plank Walk | 76.4 | 1.77 | 74.63 – 78.17 | No | Yes |
| Plank Baseline | 77.5 | 2.50 | 75.00- 80.00 | No | Yes |

| | | | | | |
|---|---|---|---|---|---|
| Plank Recovery | 80.9 | 2.55 | 78.35 – 83.45 | Yes | No |

*Note.* Overlapping of SE bars of the separate TSST event phases with the Baseline Standing condition was checked with the "Stressor Event Detection" column indicating whether the RD was able to detect a difference in Instantaneous HR when compared to the Baseline Standing condition.

**Figure 12**

*Event Difference Line Plot Plank*



*Note.* Instantaneous HR difference (Embrace-RD) is shown by event. Each participant is represented with a blue line. No a priori boundary was set as no stressor was detected by the RD.

Hypothesis 3, which stated that PPG HR data collected from the Embraceplus will not be in satisfactory agreement when compared to data collected from the gold standard ECG device when tested for the event level, was accepted.

# Discussion

This study was conducted to determine the validity of CVA data collected from the Empatica Embrace Plus PPG wearable when compared to data collected from a gold-standard ECG RD in a virtual reality stress experiment setting. The EmbracePlus was found not to display a satisfactory level of agreement with the RD in terms of average HR and SD of RR interval measurement. RMSSD data from the EmbracePlus was found to be valid compared to the RD, and average HR data was also found to agree with the RD in lower HR values (<80). The EmbracePlus was additionally found not to show satisfactory agreement with the RD in terms of detecting stressors with HR data.

## Movement-Induced Noise

The findings for the accepted hypotheses can be explained in part by the general unreliability of the PPG during movement, in line with previous research on the effects of movement on PPG wearables (Coelli et al., 2024; Mühlen et al., 2022; Menghini et al., 2019). In Figure 3, large amounts of noise can be seen with the EmbracePlus displaying considerably larger noise than the RD. This noise is the most prominent in stages of the experiment where there is movement or speech, further supporting this viewpoint. Adding to this, the parameter-level data was extracted from the entire experiment, which contained multiple instances of walking and body movement, creating noise. In the TSST condition also, the findings of the EmbracePlus not conforming to RD results (see Figure 9) are again reasoned by its PPG scanners' susceptibility to speech and wrist movement related artefacts, which create noise in RR interval data, negatively affecting validity (Mühlen et al., 202; Menghini et al., 2019). This PPG sensitivity to movement is seen more clearly in the plank condition, where the participant moved throughout the entire event. Looking at Figure 12, the difference between the devices is more prevalent than in the TSST, with double the amount of events exceeding the acceptable difference of the TSST condition. Furthermore, unlike the TSST, the trends between detected HR levels in the devices also differ (see Table 3 and 4). These factors suggest increased disagreement with events that include more movement.

## HR and RR Interval SD

Data noise could also partially explain the visible increase in the difference between the EmbracePlus and RD in HR measurements. As heart rate increases, the devices are relatively aligned until the mean HR exceeds 80 bpm, at which point the EmbracePlus begins to significantly underestimate ECG HR readings (see Figure 4). This finding aligns with a study by Jo et al. (2016), which found that a PPG device underestimates HR at high RD mean HR conditions. This could potentially be explained by the smaller gaps between RR intervals when HR is high, which exacerbates the effects of noisy RR interval data, making it increasingly difficult to detect accurate HR values with the PPG. If this is the case, it may further negatively affect the usability of the EmbracePlus in stress research as the triggering of high HR is often attempted using stressors as a core feature of experiments. However, this position is speculative and requires deeper analysis utilising manual preprocessing of the raw PPG data with sophisticated signal cleaning.

## RMSSD

Initially, the validity of the RMSSD measurements could be explained by time domain measures, including RMSSD being potentially resilient towards data artefacts caused by noise from speech and movement (Sheridan et al., 2020).  However, taking a closer look at the findings, multiple discrepancies can be seen. Firstly, the RMSSD values per device were abnormally low for the RD with a mean of only 4.78ms, while research suggests that normal average RMSSD values fall around 30-80ms depending on the source and age group (G. Kim & Woo, 2011) (Tegegne et al., 2019) (Jarczok et al., 2019). Whilst RMSSD values below 12ms are possible, they may indicate high chronic stress and cardiovascular problems, which are unlikely to exist in the entire sample, considering both the exclusion of participants with cardiovascular diseases and the results indicating a low amount of stressor detection discussed below (Srinivasan et al., 2024). Additionally, the EmbracePlus's mean RMSSD was over four times that of the RD's at 21.24 ms, but the difference values still fell within the a priori boundaries of ±70 ms. This indicates that the a priori boundaries may not have been sensitive enough to validate the device in the aforementioned extraordinary condition. While the hypothesis remains rejected following the Van Lier et al. (2019) framework, these findings should be interpreted with caution.

**Stressor Detection (Event Level)**

Analysis using the framework of Van Lier et al. (2019) only detected a significant enough change in HR from the baseline during the speech phase of the TSST. However, the general HR trend seen Figure 7 and Table 1 showcases a part of the body's ANS response to stress as heart rate increases during stressful situations, like the task instructions and speech, before decreasing once they are completed as seen in the recover condition (O'Connor et al., 2020; Dunlavey, 2018; Kim et al., 2018). The fact that the arithmetic task was not detected by the RD, even though it was designed to be stress-inducing, could potentially be explained by participants habituating to the stress effects of the speech, as the arithmetic task occurred directly after the speech portion (Tyra et al., 2021).

Additionally, even with movement related data artifacts causing the EmbracePlus to not conform to the RD, the EmbracePlus still displays a HR decrease in the TSST recovery condition compared to the stressors which conform with the RD, possibly suggesting a small degree of agreement in terms of HR trend per event marker (when mean HR increases and decreases)  (see Table 2).

It is important to note that the findings of the RD not detecting a stress response for the walk-the-plank condition could potentially be explained by participants not having height intolerances, as a study by Bzdúšková et al. (2022) showed that participants without a fear of heights showed less HR response to a similar VR condition. Another possible explanation is that the RD data accuracy may have been compromised due to tugging of the electrode cables while the participant walked around.

## Limitations and Future Research

There were multiple key limitations in this study. Firstly, there was the existence of noise and artefacts in IBI RR interval data. This could be partly viewed as a feature of the validity of the PPG, but it can also be assumed that studies that use the EmbracePlus post-validation will clean artefacts from their data sets. Because of this, validation studies for the EmbracePlus should also do so. While such quality checks and advanced data cleaning fall outside of the domain of this study, it is important for future validation research to do so. Secondly, due to the participant sample being mostly homogenous in terms of skin tone, it was not possible to research the implications of skin darkness in relation to PPG validity suggested by Menghini et al. (2019). Future research should include an ethnically diverse sample to ensure PPG validity statements are inclusive for people with dark skin tones. The third limitation was the sample size of (*N* = 18) participants in the study; previous validation studies, such as that of Van Lier et al. (2019), used sample sizes of over 50, which would theoretically decrease the chances of type 2 errors. Furthermore, previous studies using the plank condition that detected stress responses utilised much larger samples of 42-87, which could potentially explain why the stressor was not detected in this study (Basbasse et al., 2023; Martens et al., 2019).

## Strengths

The study also has notable strengths. Most notably, a strength was the usage of VR, which strengthened ecological validity in the study. This helped emulate the outside-of-laboratory conditions PPG wearables like the EmbracePlus would normally be used in, post-validation, and hence strengthened the validity of the validation itself. Moreover, VR allowed for stressor events like the plank to be used when such conditions are normally impossible to safely replicate in a laboratory. Another strength was the usage of all data collected in the experiment to conduct parameter-level analysis instead of only using data from stress-related experimental events. This allowed for a validation of data collected continuously over a long period of time (the experiment duration of about 1 hour), which can be useful for post-validation research using the EmbracePlus to study the long-term effects of stress on CVA.

## Conclusion

To conclude, this study has validated CVA data from the EmbracePlus PPG wearable in a VR stress research environment by comparing it to a gold standard ECG RD. Hence, the EmbracePlus was found to not show sufficient agreement with the RD in measuring said data in SD of RR interval form as well as in stressor detection through instantaneous HR measurement. RMSSD data collected from the EmbracePlus was, however, found to be in agreement with the RD. Furthermore, possibly due to data noise while high HR levels were measured, HR data were found to be more in agreement in lower mean HR values but became more divergent from the RD as HR increased. For researchers and clinicians, these findings suggest caution in using the EmbracePlus for CVA measurement in VR stress research. This study also showcases the importance of conducting validation research for PPG wearables to acquire the much-needed understanding of how well these devices work compared to laboratory alternatives. Future research should further validate the EmbracePlus CVA data in VR stress research environments, making use of sophisticated data cleaning as well as skin colour diverse and large samples.

# References

Ahmed, T., Qassem, M., & Kyriacou, P. A. (2023). Measuring stress: a review of the current cortisol and dehydroepiandrosterone (DHEA) measurement techniques and considerations for the future of mental health monitoring. *Stress*, *26*(1), 29–42. https://doi.org/10.1080/10253890.2022.2164187

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2016). The Trier Social Stress Test: Principles and Practice. *Neurobiology of Stress*, *6*, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001

Barnett, P. A., Spence, J. D., Manuck, S. B., & Jennings, J. R. (1997). Psychological stress and the progression of carotid artery disease. *Journal of Hypertension*, *15*(1), 49–55. https://doi.org/10.1097/00004872-199715010-00004

Basbasse, Y. E., Packheiser, J., Peterburs, J., Maymon, C., Güntürkün, O., Grimshaw, G., & Ocklenburg, S. (2023). Walk the plank! Using mobile electroencephalography to investigate emotional lateralization of immersive fear in virtual reality. *Royal Society Open Science*, *10*(5). https://doi.org/10.1098/rsos.221239

Beh, H. C. (1998). Cardiovascular reactivity to psychological stressors. *Australian Journal of Psychology*, *50*(1), 49–54. https://doi.org/10.1080/00049539808257531

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Bzdúšková, D., Marko, M., Hirjaková, Z., Kimijanová, J., Hlavačka, F., & Riečanský, I. (2022). The effects of virtual height exposure on postural control and psychophysiological stress are moderated by individual height intolerance. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.773091

CBS. (2022, March 2). 1 in 5 young workers experience work-related stress. *Statistics Netherlands*. https://www.cbs.nl/en-gb/news/2022/09/1-in-5-young-workers-experience-work-related-stress

Cho, D., Ham, J., Oh, J., Park, J., Kim, S., Lee, N., & Lee, B. (2017). Detection of Stress Levels from Biosignals Measured in Virtual Reality Environments Using a Kernel-Based Extreme Learning Machine. *Sensors*, *17*(10), 2435. https://doi.org/10.3390/s17102435

Coelli, S., Carrara, M., Bianchi, A. M., De Tommaso, M., Actis-Grosso, R., & Reali, P. (2024). Comparing consumer and Research-Grade wristbands for Inter-Beat intervals monitoring. *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 201–206. https://doi.org/10.1109/metroxraine62247.2024.10797108

Damgrave, R. (2016). Virtual reality. In *Springer eBooks* (pp. 1–3). https://doi.org/10.1007/978-3-642-35950-7_6472-4

Dunlavey, C. J. (2018, June 15). *Introduction to the Hypothalamic-Pituitary-Adrenal Axis: healthy and dysregulated stress responses, developmental stress and neurodegeneration*. https://pmc.ncbi.nlm.nih.gov/articles/PMC6057754/#:~:text=The%20hypothalamic%2Dpituitary%2Dadrenal%20axis%20(HPA)%20is%20the,Breedlove%20and%20Watson%2C%202013

*ECG Electrocardiogram Smart Amplifier*. (2025). BIOPAC. Retrieved May 9, 2025, from https://www.biopac.com/product/smart-amp-ecg/#product-tabs

*EEG/ECG/EMG/EOG Prep Gel 114 g*. (2024, August 1). BIOPAC Systems, Inc. https://www.biopac.com/product/eegecgemgeog-prep-gel-114-g/

*Electrode Gel, 250 g*. (2024, August 1). BIOPAC Systems, Inc. https://www.biopac.com/product/electrode-gel-250-g/

Elgendi, M., Jonkman, M., & De Boer, F. (2010). Frequency bands effects on QRS detection. In DBLP, *DBLP*. https://www.researchgate.net/publication/221334234

Empatica. (2025a). *EmbracePlus | The world's most advanced smartwatch for continuous health monitoring*. Retrieved May 20, 2025, from https://www.empatica.com/en-int/embraceplus

Empatica. (2025b). *Empatica Care | Empatica Health Monitoring Platform | Software*. Retrieved May 9, 2025, from https://www.empatica.com/en-int/care/

Finseth, T. T., Smith, B., Van Steenis, A. L., Glahn, D. C., Johnson, M., Ruttle, P., Shirtcliff, B. A., & Shirtcliff, E. A. (2024). When Virtual Reality becomes Psychoneuroendocrine Reality: A stress(or) review. *Psychoneuroendocrinology*, *166*, 107061. https://doi.org/10.1016/j.psyneuen.2024.107061

Gerboni, G., Comunale, G., Chen, W., Taylor, J. L., Migliorini, M., Picard, R., Cruz, M., & Regalia, G. (2023). Prospective clinical validation of the Empatica EmbracePlus wristband as a reflective pulse oximeter. *Frontiers in Digital Health*, *5*. https://doi.org/10.3389/fdgth.2023.1258915

Glise, K., Ahlborg, G., & Jonsdottir, I. H. (2014). Prevalence and course of somatic symptoms in patients with stress-related exhaustion: does sex or age matter. *BMC Psychiatry*, *14*(1). https://doi.org/10.1186/1471-244x-14-118

GmbH, I. (2025). *Cyberduck | Libre server and cloud storage browser for Mac and Windows with support for FTP, SFTP, WebDAV, Amazon S3, OpenStack Swift, Backblaze B2, Microsoft Azure & OneDrive, Google Drive and Dropbox*. Retrieved May 9, 2025, from https://cyberduck.io/

Hapke, U., Maske, U., Scheidt-Nave, C., Bode, L., Schlack, R., & Busch, M. (2013). Chronischer Stress bei Erwachsenen in Deutschland. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *56*(5–6), 749–754. https://doi.org/10.1007/s00103-013-1690-9

Hendriks, A. a. J., Ormel, J., & Van De Willige, G. (1990). Long-Term difficulties inventory [Dataset]. In *PsycTESTS Dataset*. https://doi.org/10.1037/t30127-000

*Illustration of QRS complexes and RR interval of ECG signals.* (n.d.). ResearchGate. https://www.researchgate.net/figure/llustration-of-QRS-complexes-and-RR-interval-of-ECG-signals_fig4_326588784/actions#reference

International Physical Activity Questionnaire. (1998). *International Physical Activity questionnaire - short form* [Report]. https://youthrex.com/wp-content/uploads/2019/10/IPAQ-TM.pdf

Jarczok, M. N., Koenig, J., Wittling, A., Fischer, J. E., & Thayer, J. F. (2019). First Evaluation of an Index of Low Vagally-Mediated Heart Rate variability as a marker of health risks in human adults: proof of concept. *Journal of Clinical Medicine*, *8*(11), 1940. https://doi.org/10.3390/jcm8111940

Jo, E., Lewis, K., Directo, D., Kim, M. J., & Dolezal, B. A. (2016, August 5). *Validation of biofeedback wearables for photoplethysmographic heart rate tracking*. https://pmc.ncbi.nlm.nih.gov/articles/PMC4974868/

Kim, G., & Woo, J. (2011). Determinants for heart rate variability in a normal Korean population. *Journal of Korean Medical Science*, *26*(10), 1293. https://doi.org/10.3346/jkms.2011.26.10.1293

Kim, H., Cheon, E., Bai, D., Lee, Y. H., & Koo, B. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, *15*(3), 235–245. https://doi.org/10.30773/pi.2017.08.17

Kirschbaum, C., Pirke, K., & Hellhammer, D. H. (1993). The ‘Trier Social Stress Test’ – a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1–2), 76–81. https://doi.org/10.1159/000119004

Kleckner, I. R., Feldman, M. J., Goodwin, M. S., & Quigley, K. S. (2020). Framework for selecting and benchmarking mobile devices in psychophysiological research. *Behavior Research Methods*, *53*(2), 518–535. https://doi.org/10.3758/s13428-020-01438-9

Li, Y. (2024). Impact of psychological stress on cardiovascular health and strategies for management. *Lecture Notes in Education Psychology and Public Media*, *63*(1), 113–118. https://doi.org/10.54254/2753-7048/63/20240940

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Martens, M. A., Antley, A., Freeman, D., Slater, M., Harrison, P. J., & Tunbridge, E. M. (2019). It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology*, *33*(10), 1264–1273. https://doi.org/10.1177/0269881119860156

Menghini, L., Gianfranchi, E., Cellini, N., Patron, E., Tagliabue, M., & Sarlo, M. (2019). Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology*, *56*(11). https://doi.org/10.1111/psyp.13441

Monk, E. (2023). The Monk Skin Tone Scale. *SocArXiv Papers*. https://doi.org/10.31235/osf.io/pdf4c

Mühlen, J. M., Stang, J., Skovgaard, E. L., Judice, P. B., Molina-Garcia, P., Johnston, W., Sardinha, L. B., Ortega, F. B., Caulfield, B., Bloch, W., Cheng, S., Ekelund, U., Brønd, J. C., Grøntved, A., & Schumann, M. (2021). Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE Network. *British Journal of Sports Medicine*, *55*(14), 767–779. https://doi.org/10.1136/bjsports-2020-103148

Nakao, M. (2017). Somatic manifestation of distress: clinical medicine, psychological, and public health perspectives. *BioPsychoSocial Medicine*, *11*(1). https://doi.org/10.1186/s13030-017-0119-3

O’Connor, D. B., Thayer, J. F., & Vedhara, K. (2020). Stress and Health: A review of Psychobiological processes. *Annual Review of Psychology*, *72*(1), 663–688. https://doi.org/10.1146/annurev-psych-062520-122331

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00660

*PPG signal with heartbeat rate R-R time interval.* (n.d.). ResearchGate. https://www.researchgate.net/figure/PPG-signal-with-heartbeat-rate-R-R-time-interval_fig1_323408927

Sheridan, D. C., Dehart, R., Lin, A., Sabbaj, M., & Baker, S. D. (2020). Heart rate variability analysis: How much artifact can we remove? *Psychiatry Investigation*, *17*(9), 960–965. https://doi.org/10.30773/pi.2020.0168

Srinivasan, A. G., Smith, S. S., Pattinson, C. L., Mann, D., Sullivan, K., Salmon, P., & Soleimanloo, S. S. (2024). Heart rate variability as an indicator of fatigue: A structural equation model approach. *Transportation Research Part F Traffic Psychology and Behaviour*, *103*, 420–429. https://doi.org/10.1016/j.trf.2024.04.015

Steptoe, A., & Kivimäki, M. (2012). Stress and cardiovascular disease. *Nature Reviews Cardiology*, *9*(6), 360–370. https://doi.org/10.1038/nrcardio.2012.45

Tegegne, B. S., Man, T., Van Roon, A. M., Snieder, H., & Riese, H. (2019). Reference values of heart rate variability from 10-second resting electrocardiograms: the Lifelines Cohort Study. *European Journal of Preventive Cardiology*, *27*(19), 2191–2194. https://doi.org/10.1177/2047487319872567

Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2011). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, *36*(2), 747–756. https://doi.org/10.1016/j.neubiorev.2011.11.009

Tyra, A. T., Cook, T. E., Young, D. A., Hurley, P. E., Oosterhoff, B. J., John-Henderson, N. A., & Ginty, A. T. (2021). Adverse childhood experiences, sex, and cardiovascular habituation to repeated stress. *Biological Psychology*, *165*, 108175. https://doi.org/10.1016/j.biopsycho.2021.108175

Udhayakumar, R., Rahman, S., Buxi, D., Macefield, V. G., Dawood, T., Mellor, N., & Karmakar, C. (2023). Measurement of stress-induced sympathetic nervous activity using multi-wavelength PPG. *Royal Society Open Science*, *10*(8). https://doi.org/10.1098/rsos.221382

Van Lier, H. G., Pieterse, M. E., Garde, A., Postel, M. G., De Haan, H. A., Vollenbroek-Hutten, M. M. R., Schraagen, J. M., & Noordzij, M. L. (2019). A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor. *Behavior Research Methods*, *52*(2), 607–629. https://doi.org/10.3758/s13428-019-01263-9

*VR Game Development Software & Engine | Unity*. (n.d.). Unity. https://unity.com/solutions/vr

WHO. (2022, June 17). *Stress*. https://www.who.int/news-room/questions-and-answers/item/stress

## Appendix A
## R Studio Code

### Parameter pre process

```
data <- read.csv("C:/Users/alpyu/Desktop/dataparameter/Part03_aligned_rrandhr_200Hz_withPhases.csv")  # Replace with your actual filename or path

# Define RMSSD function (converts RR from seconds to milliseconds)

rmssd <- function(rr) {
  rr <- na.omit(rr)        # Remove NA values
  rr_ms <- rr * 1000       # Convert RR intervals from seconds to milliseconds
  diff_rr <- diff(rr_ms)    # Calculate successive differences
  sqrt(mean(diff_rr^2))     # Calculate RMSSD
}
# Calculate RMSSD
rmssd_biopac <- rmssd(data$RR_Int_Biopac)
rmssd_embrace <- rmssd(data$RR_Int_Embrace)

# Calculate Mean HR
mean_HR_biopac <- mean(data$HR_Biopac_1min, na.rm = TRUE)
mean_HR_embrace <- mean(data$HR_Embrace_1min, na.rm = TRUE)

# Calculate Standard Deviation of RR Intervals (in ms)
sd_RR_biopac <- sd(data$RR_Int_Biopac * 1000, na.rm = TRUE)   # Convert to ms
sd_RR_embrace <- sd(data$RR_Int_Embrace * 1000, na.rm = TRUE)  # Convert to ms

# Print Results
cat("Reference Device (Biopac):\n")
cat("Mean HR (bpm):", round(mean_HR_biopac, 2), "\n")
cat("SD of RR Intervals (ms):", round(sd_RR_biopac, 2), "\n")
cat("RMSSD (ms):", round(rmssd_biopac, 2), "\n\n")

cat("Watch Device (Embrace):\n")
cat("Mean HR (bpm):", round(mean_HR_embrace, 2), "\n")
cat("SD of RR Intervals (ms):", round(sd_RR_embrace, 2), "\n")
cat("RMSSD (ms):", round(rmssd_embrace, 2), "\n")
```

## Parameter Plots

## HR

```
# Load libraries
library(readxl)
library(ggplot2)

# Step 1: Load the Excel file
data <- read_excel("C:/Users/alpyu/Desktop/paramter processeesd.xlsx")

# Step 2: Compute mean and difference between devices
data$HR_mean <- (data$Av_HR_Embrace + data$Av_HR_RD) / 2
data$HR_diff <- data$Av_HR_Embrace - data$Av_HR_RD

# Step 3: Compute Bland-Altman statistics
mean_diff <- mean(data$HR_diff, na.rm = TRUE)
sd_diff <- sd(data$HR_diff, na.rm = TRUE)
upper_limit <- mean_diff + 1.96 * sd_diff
lower_limit <- mean_diff - 1.96 * sd_diff

# Step 4: Create the Bland-Altman plot
ggplot(data, aes(x = HR_mean, y = HR_diff)) +
  geom_point(shape = 21, fill = "skyblue", color = "black", size = 3) +
  geom_hline(yintercept = mean_diff, color = "blue", linetype = "solid", size = 1) +
  geom_hline(yintercept = upper_limit, color = "red", linetype = "dashed", size = 1) +
  geom_hline(yintercept = lower_limit, color = "red", linetype = "dashed", size = 1) +
  geom_hline(yintercept = 5, color = "darkgreen", linetype = "dotdash", size = 0.8) +
  geom_hline(yintercept = -5, color = "darkgreen", linetype = "dotdash", size = 0.8) +
  geom_hline(yintercept = 0, color = "black", linetype = "dotted", size = 0.8) +
  labs(
    title = "HR",
    x = "Mean HR (Embrace,RD)",
    y = "Difference HR (Embrace - RD)"
  ) +
  theme_minimal() +
  annotate("text", x = max(data$HR_mean, na.rm = TRUE), y = mean_diff,
       label = paste0("Mean = ", round(mean_diff, 2)), hjust = 1.1) +
  annotate("text", x = max(data$HR_mean, na.rm = TRUE), y = upper_limit,
       label = paste0("+1.96SD = ", round(upper_limit, 2)), hjust = 1.1) +
  annotate("text", x = max(data$HR_mean, na.rm = TRUE), y = lower_limit,
       label = paste0("-1.96SD = ", round(lower_limit, 2)), hjust = 1.1)
```

## **SD of RR/PP,**

```
# Load required libraries
library(readxl)
library(ggplot2)

# Load Excel file
data <- read_excel("C:/Users/alpyu/Desktop/paramter processeesd.xlsx")

# Convert RR from ms to seconds
data$SD_RR_Embrace <- data$SD_RR_Embrace / 1000
data$SD_RR_RD <- data$SD_RR_RD / 1000

# Calculate mean and difference
data$SDRR_mean <- (data$SD_RR_Embrace + data$SD_RR_RD) / 2
data$SDRR_diff <- data$SD_RR_Embrace - data$SD_RR_RD

# Bland-Altman stats
mean_diff <- mean(data$SDRR_diff, na.rm = TRUE)
```

```
sd_diff <- sd(data$SDRR_diff, na.rm = TRUE)

upper_limit <- mean_diff + 1.96 * sd_diff

lower_limit <- mean_diff - 1.96 * sd_diff


# y-axis includes green boundaries ±0.06 and BA limits

y_min <- min(-0.06, lower_limit, min(data$SDRR_diff, na.rm = TRUE)) - 0.01

y_max <- max(0.06, upper_limit, max(data$SDRR_diff, na.rm = TRUE)) + 0.01


# Create plot

ggplot(data, aes(x = SDRR_mean, y = SDRR_diff)) +

  geom_point(shape = 21, fill = "skyblue", color = "black", size = 3) +

  geom_hline(yintercept = 0, color = "black", linetype = "dotted", size = 0.8) +

  geom_hline(yintercept = mean_diff, color = "blue", linetype = "solid", size = 1) +

  geom_hline(yintercept = upper_limit, color = "red", linetype = "dashed", size = 1) +

  geom_hline(yintercept = lower_limit, color = "red", linetype = "dashed", size = 1) +

  geom_hline(yintercept = 0.06, color = "darkgreen", linetype = "dotdash", size = 0.8) +

  geom_hline(yintercept = -0.06, color = "darkgreen", linetype = "dotdash", size = 0.8) +

  coord_cartesian(ylim = c(y_min, y_max)) +

  labs(

    title = "SD of RR",

    x = "Mean SD_RR (Embrace,RD)",

    y = "Difference SD_RR (Embrace - RD)"

  ) +

  theme_minimal() +

  annotate("text", x = max(data$SDRR_mean, na.rm = TRUE), y = mean_diff,

           label = paste0("Mean = ", round(mean_diff, 3)), hjust = 1.1) +

  annotate("text", x = max(data$SDRR_mean, na.rm = TRUE), y = upper_limit,

           label = paste0("+1.96SD = ", round(upper_limit, 3)), hjust = 1.1) +

  annotate("text", x = max(data$SDRR_mean, na.rm = TRUE), y = lower_limit,
           label = paste0("-1.96SD = ", round(lower_limit, 3)), hjust = 1.1) +
  annotate("text", x = min(data$SDRR_mean, na.rm = TRUE), y = 0.06,
           label = "+0.06s", vjust = -1, hjust = 0, color = "darkgreen") +
  annotate("text", x = min(data$SDRR_mean, na.rm = TRUE), y = -0.06,
           label = "-0.06s", vjust = 1.5, hjust = 0, color = "darkgreen")
```

**RMSSD**

```
# Load libraries
library(readxl)
library(ggplot2)

# Load data
data <- read_excel("C:/Users/alpyu/Desktop/paramter processeesd.xlsx")

# Convert RMSSD values from ms to seconds
data$RMSSD_Embrace <- data$RMSSD_Embrace / 1000
data$RMSSD_RD <- data$RMSSD_RD / 1000

# Calculate mean and difference
data$RMSSD_mean <- (data$RMSSD_Embrace + data$RMSSD_RD) / 2
data$RMSSD_diff <- data$RMSSD_Embrace - data$RMSSD_RD

# Bland–Altman stats
mean_diff <- mean(data$RMSSD_diff, na.rm = TRUE)
sd_diff <- sd(data$RMSSD_diff, na.rm = TRUE)
upper_limit <- mean_diff + 1.96 * sd_diff
lower_limit <- mean_diff - 1.96 * sd_diff

# ±0.07s lines
```

```
y_min <- min(-0.07, lower_limit, min(data$RMSSD_diff, na.rm = TRUE)) - 0.01
y_max <- max(0.07, upper_limit, max(data$RMSSD_diff, na.rm = TRUE)) + 0.01

# Plot
ggplot(data, aes(x = RMSSD_mean, y = RMSSD_diff)) +
  geom_point(shape = 21, fill = "skyblue", color = "black", size = 3) +
  geom_hline(yintercept = 0, color = "black", linetype = "dotted", size = 0.8) +
  geom_hline(yintercept = mean_diff, color = "blue", linetype = "solid", size = 1) +
  geom_hline(yintercept = upper_limit, color = "red", linetype = "dashed", size = 1) +
  geom_hline(yintercept = lower_limit, color = "red", linetype = "dashed", size = 1) +
  geom_hline(yintercept = 0.07, color = "darkgreen", linetype = "dotdash", size = 0.8) +
  geom_hline(yintercept = -0.07, color = "darkgreen", linetype = "dotdash", size = 0.8) +
  coord_cartesian(ylim = c(y_min, y_max)) +
  labs(
    title = "RMSSD",
    x = "Mean RMSSD (Embrace, RD)",
    y = "Difference RMSSD (Embrace - RD)"
  ) +
  theme_minimal() +
  annotate("text", x = max(data$RMSSD_mean, na.rm = TRUE), y = mean_diff,
        label = paste0("Mean = ", round(mean_diff, 3)), hjust = 1.1) +
  annotate("text", x = max(data$RMSSD_mean, na.rm = TRUE), y = upper_limit,
        label = paste0("+1.96SD = ", round(upper_limit, 3)), hjust = 1.1) +
  annotate("text", x = max(data$RMSSD_mean, na.rm = TRUE), y = lower_limit,
        label = paste0("-1.96SD = ", round(lower_limit, 3)), hjust = 1.1) +
  annotate("text", x = min(data$RMSSD_mean, na.rm = TRUE), y = 0.07,
        label = "+0.07s", vjust = -1, hjust = 0, color = "darkgreen") +
  annotate("text", x = min(data$RMSSD_mean, na.rm = TRUE), y = -0.07,
        label = "-0.07s", vjust = 1.5, hjust = 0, color = "darkgreen")
```

## EVENT LEVEL POST PROCESSING

### SST

```
# Load libraries
library(tidyverse)

# Set folder path containing the participant CSV files
folder_path <- "path/to/your/csv/files"  # 🔁 Replace with your actual folder path

# Read all CSV files and add Participant ID based on file name
all_data <- list.files(path = folder_path, pattern = "\\.csv$", full.names = TRUE) %>%
 map_df(~ {
  df <- read_csv(.x, show_col_types = FALSE)
  df$Participant <- tools::file_path_sans_ext(basename(.x))
  return(df)
 })

# Define the event marker
target_events <- c(
 "Baseline Seated",
 "TSST Instructions",
 "Anticipation Phase",
 "Speech Phase",
 "Mental Arithmetic"
)

# Filter to only target events and reshape HR variables to long format
filtered_data <- all_data %>%
 filter(Event_Phases %in% target_events) %>%
 pivot_longer(
  cols = c(HR_Biopac_Smooth, HR_Embrace_Smooth),
  names_to = "Device",
  values_to = "HR"
 ) %>%
 mutate(
  Device = recode(Device,
                 "HR_Biopac_Smooth" = "Biopac",
                 "HR_Embrace_Smooth" = "Embrace")
 )

# Compute average HR per participant, per device, per event
avg_hr <- filtered_data %>%
 group_by(Participant, Device, Event_Phases) %>%
 summarise(
  Average_HR = mean(HR, na.rm = TRUE),
  .groups = "drop"
 )

# View result
print(avg_hr)

# Save the result to a CSV file
write_csv(avg_hr, "average_hr_by_participant_device_event.csv")

library(tidyverse)

# Load your existing CSV with the long format (device column)
avg_hr_long <- read_csv("average_hr_by_participant_device_event.csv", show_col_types = FALSE)
```

```
# Pivot wider to create separate columns for Biopac and Embraceplus
avg_hr_wide <- avg_hr_long %>%
  pivot_wider(
    names_from = Device,
    values_from = Average_HR
  )

# Check the result
print(avg_hr_wide)

# Save the new wide-format CSV
write_csv(avg_hr_wide, "average_hr_wide_by_participant_event.csv")
```

### embrace SST

```
library(tidyverse)

# Read data
avg_hr_wide <- read_csv("average_hr_wide_by_participant_event.csv", show_col_types = FALSE)

# Reshape to long
avg_hr_long <- avg_hr_wide %>%
  pivot_longer(cols = c(Biopac, Embrace), names_to = "Device", values_to = "Average_HR")

# Order events
event_order <- c("Baseline Seated", "TSST Instructions", "Anticipation Phase",
                 "Speech Phase", "Mental Arithmetic", "TSST Recovery")
```

```
avg_hr_long <- avg_hr_long %>%
 mutate(Event_Phases = factor(Event_Phases, levels = event_order))

# Filter for Embrace only
embrace_data <- avg_hr_long %>% filter(Device == "Embrace")

# Summary: mean and SE
embrace_summary <- embrace_data %>%
 group_by(Event_Phases) %>%
 summarise(
  mean_HR = mean(Average_HR, na.rm = TRUE),
  se_HR = sd(Average_HR, na.rm = TRUE) / sqrt(n()),
  .groups = "drop"
 )

# Plot with SE lines
ggplot() +
 # Participant lines
 geom_line(data = embrace_data, aes(x = Event_Phases, y = Average_HR, group = Participant),
      color = "blue", alpha = 0.4) +
 # SE error bars
 geom_errorbar(data = embrace_summary,
        aes(x = Event_Phases, ymin = mean_HR - se_HR, ymax = mean_HR + se_HR),
        width = 0.2, color = "black") +
 # Mean line in red
 geom_line(data = embrace_summary,
      aes(x = Event_Phases, y = mean_HR, group = 1),
      color = "red", size = 1.2) +
 labs(title = "Mean HR by TSST Event - Embrace",
   x = "Event Phase",
   y = "Average HR (BPM)") +
 theme_minimal() +
 theme(axis.text.x = element_text(angle = 45, hjust = 1))
print(embrace_summary)
```

**RD SST**

```
library(tidyverse)

# Read data
avg_hr_wide <- read_csv("average_hr_wide_by_participant_event.csv", show_col_types = FALSE)

# Reshape to long
avg_hr_long <- avg_hr_wide %>%
  pivot_longer(cols = c(Biopac, Embrace), names_to = "Device", values_to = "Average_HR")

# Order events
event_order <- c("Baseline Seated", "TSST Instructions", "Anticipation Phase",
                 "Speech Phase", "Mental Arithmetic", "TSST Recovery")

avg_hr_long <- avg_hr_long %>%
  mutate(Event_Phases = factor(Event_Phases, levels = event_order))

# Filter for RD (Biopac)
rd_data <- avg_hr_long %>% filter(Device == "Biopac")

# Summary: mean and SE
rd_summary <- rd_data %>%
  group_by(Event_Phases) %>%
  summarise(
    mean_HR = mean(Average_HR, na.rm = TRUE),
    se_HR = sd(Average_HR, na.rm = TRUE) / sqrt(n()),
    .groups = "drop"
  )

# Plot with SE lines (error bars instead of ribbon)
ggplot() +
  # Participant lines
  geom_line(data = rd_data, aes(x = Event_Phases, y = Average_HR, group = Participant),
            color = "blue", alpha = 0.4) +
  # SE error bars
  geom_errorbar(data = rd_summary,
                aes(x = Event_Phases, ymin = mean_HR - se_HR, ymax = mean_HR + se_HR),
                width = 0.2, color = "black") +
  # Mean line in red
  geom_line(data = rd_summary,
            aes(x = Event_Phases, y = mean_HR, group = 1),
            color = "red", size = 1.2) +
  labs(title = "Mean HR by TSST Event - RD",
       x = "Event Phase",
       y = "Average HR (BPM)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(rd_summary)
```

### **Combination(Differences event level)**

```
library(tidyverse)
```

```
# Step 1: Read wide-format data
avg_hr_wide <- read_csv("average_hr_wide_by_participant_event.csv", show_col_types = FALSE)

# Step 2: Compute HR difference
hr_diff <- avg_hr_wide %>%
  mutate(HR_Diff = Embrace - Biopac)

# Step 3: Set event order
event_order <- c("Baseline Seated", "TSST Instructions", "Anticipation Phase",
                "Speech Phase", "Mental Arithmetic", "TSST Recovery")

hr_diff <- hr_diff %>%
  mutate(Event_Phases = factor(Event_Phases, levels = event_order))

# Step 4: Compute summary stats
diff_summary <- hr_diff %>%
  group_by(Event_Phases) %>%
  summarise(
    mean_diff = mean(HR_Diff, na.rm = TRUE),
    se_diff = sd(HR_Diff, na.rm = TRUE) / sqrt(n()),
    sd_diff = sd(HR_Diff, na.rm = TRUE),
    .groups = "drop"
  )

ggplot() +
  # participant lines
  geom_line(data = hr_diff, aes(x = Event_Phases, y = HR_Diff, group = Participant),
            color = "blue", size = 0.3, alpha = 0.25) +
  # SD BARS
  geom_errorbar(data = diff_summary,
            aes(x = Event_Phases, ymin = mean_diff - sd_diff, ymax = mean_diff + sd_diff),
            width = 0.2, color = "black", size = 0.6) +

  # SE ribbon
  geom_ribbon(data = diff_summary,
           aes(x = Event_Phases, ymin = mean_diff - se_diff, ymax = mean_diff + se_diff),
           fill = "gray70", alpha = 0.3) +

  # Mean line in red
  geom_line(data = diff_summary,
            aes(x = Event_Phases, y = mean_diff, group = 1),
            color = "red", size = 1.2) +

  # +8 and -8 lines
  geom_hline(yintercept = 8, color = "lightgreen", linetype = "dashed", size = 1) +
  geom_hline(yintercept = -8, color = "lightgreen", linetype = "dashed", size = 1) +

  labs(title = "HR Difference (Embrace - RD) TSST",
     x = "Event Phase",
     y = "HR Difference (BPM)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(head(hr_diff))
```