Validating the EmbracePlus Smartwatch for measuring Electrodermal Activity and Stress Assessment in Virtual Reality: The Moderating Role of Presence

David Calandra

Master Thesis

Faculty of Behavioural, Management, and Social Sciences

Department of Psychology, Health & Technology

University of Twente

First supervisor Prof. Dr. Matthijs Noordzij

Second supervisor Magdalena Sikora M.Sc.

19th of August 2025

Abstract

Smartwatches are increasingly used for stress monitoring, but their validation in immersive and ecologically valid settings remains limited. This study evaluated the EmbracePlus smartwatch's ability to measure electrodermal activity, in particular the skin conductance level during three virtual reality stress tasks and examined whether the subjective sense of presence influences device agreement. A within-subjects design was employed with 20 healthy adults. Skin conductance level was measured using both the EmbracePlus and a reference device. A modified validation protocol was applied, assessing signal-level agreement through cross-correlation and parameter-level agreement using Bland and Altman analysis. Results indicated weak correlations between devices, with the EmbracePlus underestimating skin conductance level compared to the reference device. While both devices detected increased arousal during stress phases, presence did not significantly predict physiological changes or improve agreement. These findings emphasize the need for improved accuracy in wearable stress sensors and greater consistency in validation practices. Importantly, this study successfully applied a validation protocol developed by the broader Stress in Action project team, providing preliminary evidence that immersive environments can be used to systematically evaluate wearable stress sensors under ecologically valid conditions.

Keywords: electrodermal activity, smartwatch validation, virtual reality, presence, physiological measurement

Validating the EmbracePlus Smartwatch for Stress Assessment in Virtual Reality: The Moderating Role of Presence

Consumer smartwatches are gaining recognition for their potential in stress recording, but scientific validation of these devices remains limited, particularly in contexts that reflect the complexity of real-world environments. Currently, many existing validation studies have focused on resting-state measurements or simple stress-inducing tasks in controlled laboratory settings (Canali et al., 2022). Furthermore, while tasks like the Trier Social Stress Test (TSST) offer robust, ecologically valid stress induction, they are logistically demanding, resource-intensive, and difficult to standardize across studies (Allen et al., 2017). In such settings, the user's sense of presence, or how involved or connected they feel to the environment, is typically not examined, as it is generally considered irrelevant unless the setting involves an immersive virtual reality experience. (Peake et al., 2018; Canali et al., 2022). This presents a challenge for the validation, because smartwatches aren't just worn when someone is resting. They're frequently used in dynamic, everyday situations where stress responses are affected by both environmental factors and psychological processes. To address this issue, research requires tools that can replicate real-life complexity without compromising experimental control. Here, virtual reality (VR) emerges as a promising solution, enabling the simulation of naturalistic stress tasks with both ecological validity and systematic control (Lu & Lau, 2025; Shakila Meshkat et al., 2024). Equally important, a critical aspect accompanied by VR is the psychological concept of presence which is the subjective experience of truly “being there” in the virtual space (Triberti et al., 2025). This immersion during VR goes beyond just what we see and hear it leads to stronger physical reactions by activating the body's automatic nervous system more than in less immersive environments do (Meehan et al., 2002; Kühne et al., 2023). One common way to measure this physiological response is through electrodermal activity (EDA), a widely used, non-invasive measure of sympathetic nervous system activity and stress, reflecting sweat gland-induced changes in skin conductance (Boucsein, 2012).While valuable for stress assessment, EDA signals can be sensitive to various contextual influences, including task demands, emotional engagement, and technical factors such as sensor placement (McDuff et al., 2024). Considering the link between presence and physical response, it might be possible that presence alters EDA activity during VR (Fromberger et al., 2015; Toczek, 2018). This raises an important question: if presence affects the physiological state of stress, could it also influence the accuracy of wearable devices designed to detect these changes? Investigating this relationship is necessary to improve wearable validation efforts in contexts where VR is used.

The Challenge of Wearable Validation in VR

Although we know that VR generates realistic stressors, it has not been widely used to validate smartwatches, particularly the physiological measurement of EDA. A systematic review by Halbig and Latoschik (2021) analysing 1,119 VR studies incorporating physiological data found that only around 3% focused on sensor validation, with the majority emphasizing experiential state classification, rather than validating the sensors themselves. This suggests that while the integration of physiological monitoring in VR is expanding, few studies truly assess the accuracy of wearables. Hence, existing validation efforts often remain preliminary or employ simplified tasks that do not reflect daily life settings (Palombi et al., 2023). On top of that, Peake et al. (2018) found that only 5% of wearable devices have been formally validated against gold-standard physiological systems and further concerns have also been raised about the data quality, algorithms, and potential biases in wearable measurements (Canali et al., 2022). These critiques emphasize the need for standardized and transparent validation protocols particularly for VR.

Without such protocols, data collected in VR environments may fail to meet the reliability standards expected in laboratory research. Addressing this gap by systematically examining how presence influences wearable EDA validity represents a step forward in ensuring accurate and meaningful stress recording in immersive contexts. But what exactly is Electrodermal Activity (EDA) and why is it considered a reliable measure of stress?

Electrodermal Activity (EDA): A Pure Measure of SNS Activation

The autonomic nervous system (ANS) is part of the body's stress response, controlling physiological changes through two primary branches: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS activates the body during perceived threats with the "fight-or-flight" response, while the PNS facilitates recovery and counterbalance through the "rest-and-digest" response (Mehl et al., 2023). To measure ANS activity, physiological indicators such as electrodermal activity (EDA) are commonly used in stress research. Thus, EDA serves as a direct representation of sympathetic arousal (Laborde et al., 2017; Boucsein, 2012). This study focuses on the tonic component, the skin conductance level (SCL), which is the slower, ongoing part of EDA (Braithwaite et al., 2013). Unlike skin conductance responses (SCRs), which are quick spikes linked to short-term reactions, SCL changes more gradually and is more useful for studying continuous stress. Most commonly, reference devises (RD) measure EDA with wet electrodes placed on the fingers or palms, where many sweat glands exist (Boucsein, 2012). In contrast, smartwatches use dry electrodes on the wrist, which have fewer sweat glands and are more affected by movement, making the data less consistent and sometimes harder to interpret (van Dooren et al., 2012). To specify, RD´s provide sensitivity to subtle physiological changes and minimizing artifacts (Menghini et al., 2019; Kim et al., 2021). Lastly, EDA is a non-invasive and low-cost method for measuring physiological arousal and stress reactions in a variety of settings. Having established the physiological measure for stress, the next step is to define the key psychological construct under investigation in this study: presence.

Presence: A Multidimensional Psychological Construct

Presence refers to the psychological construct describing the subjective feeling of "being there" in a mediated or simulated environment (Triberti et al., 2025) and is often conceptualized as a multidimensional experience (Bareišytė et al., 2024). These dimensions include physical presence, which refers to the sensation of being physically within a virtual space, and social presence, which refers to meaningful interaction with virtual agents or avatars; and self-presence, which refers to the perception of one's virtual self as an extension of the real self (Bareišytė et al., 2024). In this study, only physical and social presence are considered relevant, as the experimental design did not include self-representational aspects such as hands or feet to engage in self-presence during VR.

Notably, greater physical presence might enhance or decrease the realism of the tasks, while stronger social presence may increase the perceived evaluative threat, both of which might heighten sympathetic nervous system (SNS) reactivity (Fromberger et al., 2015). A recent review by Wang et al. (2025) analysing 94 studies found that researchers frequently use physiological signals such as head movements, electromyography (EMG), and electrocardiography (ECG) to assess presence in virtual reality. However, a major problem is that these same signals are also used to measure stress and mental effort. This means that none of them can be considered unique or specific to measuring presence alone (Wang et al., 2025). Additionally, studies have shown that a user's sense of presence in VR alters physiological responses associated with autonomic arousal (Grassini & Laumann, 2020; Wang et al., 2025). Higher presence ratings have been linked to larger electrodermal responses during immersive tasks and higher levels of skin conductance. For instance, research on VR gaming has shown that playing games in VR, as opposed to traditional screen-based formats, leads to stronger feelings of presence, lower heart rate variability, and heightened subjective fear (Lemmens et al., 2021). While such studies often compare VR to non-VR conditions, less is known about how varying levels of presence within VR experiences impact physiological responses or the agreement between different measurement devices. Another interesting example of measuring presence is illustrated by the work of Eftekharifar et al. (2021), which compared restorative effects in VR following stress tasks. The study contrasted visual presence through immersion in a virtual forest, with pictorial presence, where participants viewed an image of the same forest. Results showed that the visually immersive condition led to a significantly greater reduction in electrodermal activity (EDA) compared to the pictorial condition, suggesting that physiological responses can be influenced by the level of presence. Nevertheless, this outcome was based on an experimental manipulation of the environment, rather than an investigation of individual differences in perceived presence. While the study provides useful evidence that manipulating presence can influence autonomic nervous system responses, it does not address how subjective experiences of presence relate to physiological measures. This distinction is important, as it remains uncertain whether individuals who report higher levels of presence within the same VR environment also display distinct physiological patterns, or whether such associations are mainly driven by external manipulations of the virtual setting (Siavash Eftekharifar et al., 2021). To address this critical uncertainty and ensure the reliability of physiological data collected in such contexts, it is essential to employ a systematic validation framework.

Validation Protocols for Physiological Measurement Devices

The present study adopts a modified version of the existing validation framework developed by van Lier et al. (2020), which distinguishes between three levels of analysis: signal, parameter, and event. The signal level entails comparing raw physiological signals over time, the parameter level focuses on agreement in extracted summary measures, and the event level examines whether devices reflect physiological changes across specific task phases. These levels offer a structured approach to evaluating the validity of physiological sensors. In the context of this study, which explores electrodermal activity (EDA) measured by the EmbracePlus smartwatch during immersive VR stress tasks, the signal level is relevant for assessing the device's ability to track continuous physiological changes in real time. The parameter level is useful for evaluating whether the device captures meaningful changes in skin conductance level across different experimental conditions. Although the event level offers additional insight into task-related physiological responses, the current analysis focuses on signal- and parameter-level validation. Next to this, the influence of subjective presence on physiological responses and device agreement will be examined.

The Role of Presence in Device Validation

Despite extensive research on wearable technology validation and specific investigations into presence effects on physiological responses, the potential moderating role of presence in device validation during VR remains unexplored, which is logical given the novel nature of VR validation of wearables. This study will address three interconnected challenges: (1) the need for efficient validation methods that are adaptable to frequent device developments, (2) the unexplored potential of VR to create standardised yet immersive stress induction protocols, and (3) the understudied influence of sense of presence, the subjective feeling of "being there" in a mediated environment, on both stress reactivity and measure validity.

Research Objectives and Questions

This study aim is to compare the physiological stress measurements of SCL obtained from the EmbracePlus smartwatch's sensor against those from laboratory-grade equipment to evaluate its validity, while investigating the potential moderating role of presence. By addressing both technical validation concerns and psychological influences on measurement, this research project bridges an important gap in current validation research and tries to answer the following research questions: First, to what extent do EDA signals collected with the EmbracePlus smartwatch correspond with those obtained from a validated reference device? This overarching validity question is examined at two analytical levels, focusing on raw signal agreement and the comparability of skin conductance level (SCL) as a key parameter. Second, does the subjective sense of presence in the virtual reality environment predict differences in skin conductance level (SCL) reactivity between two task contrasts: (1) a standing baseline phase and a height-induced stress phase, and (2) a seated baseline phase and a socially induced stress phase involving a speech task? Third, does presence also moderate the level of agreement between the smartwatch and the reference device?

Methods

Design

This study used a within-subjects experimental design to validate the EmbracePlus smartwatch against the gold-standard system for measuring electrodermal activity (EDA) during stress-inducing virtual reality tasks, including a social-evaluative speech task, a height-based plank walk, and a nature-themed cognitive task. The validation followed a modified version of an existing protocol, incorporating analyses at the signal level and parameter level, to evaluate the concordance between devices (van Lier et al., 2020). Additionally, the study examined how the sense of presence within virtual environments moderates the validity of wearable stress measurements. This was assessed using standardized, validated instruments targeting both presence dimensions: the Self-Presence Questionnaire (Makransky et al., 2017; Ratan & Hasler, 2009) and the Social Presence Scale (Makransky et al., 2017; Nowak & Biocca, 2003). These questionnaires have been systematically reviewed in the context of virtual reality evaluation, as highlighted in the systematic scoping review by Bareišytė et al. (2024)

Participants

Sample Size and Recruitment

This study involved a sample of twenty healthy adult, recruited verbally at the University of Twente as well as through the university's SONA research participation system. All participants provided informed consent before beginning the study, and all procedures were conducted in accordance with ethical guidelines. Prospective participants received an email outlining the study details, including the scheduled date and time of their session, a unique participant number, and links to a participant information sheet, consent form, and home questionnaires for demographic and health-related data. They were requested to review the information sheet and complete the forms prior to their lab visit. The email also provided contact information for any questions or concerns. The demographic and physical characteristics of the participants are presented in Table 1 below.

Table 1

Demographic and Physical Characteristics of the Twenty Participants

Characteristic	N	%	M (Sd)	Min – Max
Age (years)	20		23.40 (2.30)	20 – 29
Height (cm)	20		175.05 (8.38)	161.0 – 191.0
Weight (kg)	20		76.19 (19.38)	49.9 – 108.2
Waist circumference (cm)	20		81.00 (16.28)	61.0 – 122.0
Skin tone (scale 1–5)	20		3.00 (0.71)	2.0 – 5.0
Gender				
Female	12	60.0		
Male	8	40.0		
Handedness				
Left	1	5.0		
Right	19	95.0		

Note. Physical characteristics are reported as means (M), standard deviations (SD), and range (Min–Max). Gender and handedness are reported as frequency (N) and percentage (%).

Inclusion Criteria and Exclusion Criteria

Inclusion Criteria

To participate in the study, individuals were required to meet several eligibility conditions. First, participants had to be healthy adults who could provide informed consent. This criterion ensured that participants fully understood the nature of the study and their rights before participating. Additionally, participants needed to have sufficient English language proficiency to be able to follow and complete the tasks, as well as to complete all required questionnaires. Only those who provided written informed consent prior to the start of any study-related procedures were permitted to continue in the study. Finally, participants had to be generally healthy and free from any conditions that could compromise the quality or reliability of physiological data, particularly electrodermal activity and other autonomic measurements.

Exclusion Criteria

Several exclusion criteria were outlined to prevent data contamination and protect participant safety. Individuals with diagnosed heart disease were excluded, as cardiovascular conditions can significantly alter heart rate variability and electrocardiogram (ECG) signals, thereby interfering with the study's physiological measurements. The use of medications that influence the autonomic nervous system were also parts for exclusion. They were part of it, because if participants are taking beta-blockers, these medications can alter heart rate and blood pressure by blocking beta-adrenergic receptors, so they were not eligible if taken. Similarly, individuals using anti-sweating medications, such as acetylcholine blockers that affect electrodermal activity were excluded. Other disqualifying medications included heart rhythm stabilizers like ivabradine, as well as various types of antidepressants, including tricyclic antidepressants (TCAs), serotonin-norepinephrine reuptake inhibitors (SNRIs), and selective serotonin reuptake inhibitors (SSRIs), all of which can influence autonomic functioning.

Materials

Wearable Devices

The study used multiple wearable devices for physiological monitoring. The EmbracePlus smartwatch was used to measure electrodermal activity and was worn on the participant’s non-dominant hand. In addition, participants wore an Oura ring, preferably on the index finger of the same hand as the smartwatch, to collect complementary physiological data. A third device, the NOWATCH, was placed on the wrist opposite to the one wearing the

EmbracePlus, ensuring balanced placement and minimizing interference between devices. Notably, for the current paper only the EmbracePlus was analysed.

Gold Standard Equipment

Physiological measurements were continuously recorded using a BIOPAC MP160 data acquisition system (BIOPAC Systems, Inc., 2024). The system captured electrodermal activity (EDA) using disposable Ag/AgCl electrodes, which were attached to the medial phalanges of the middle- and ring fingers on the participant's non-dominant hand. Additionally, a 3-axis accelerometer was used to monitor physical movement throughout the experimental tasks, ensuring that motion-related artifacts could be identified and accounted during pre-processing.

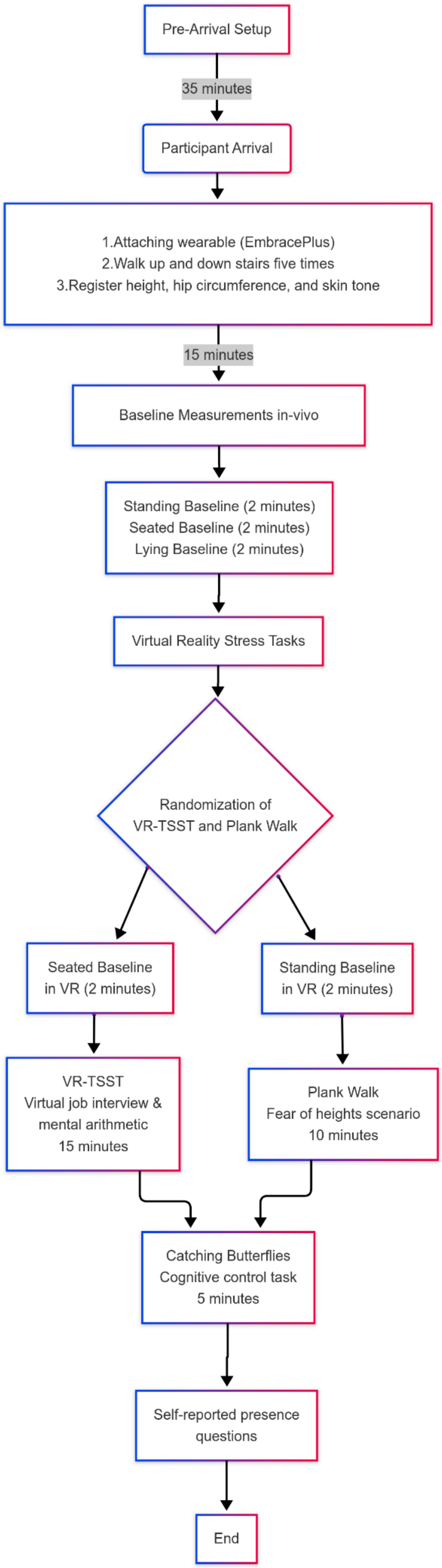
Virtual Reality Setup

The VR environment in this study was developed with a Unity software as part of the Stress in Action project. Much credit goes to M. Sikora and L. Frösler for their significant contributions to this development. Further details about the project and the VR environment will be made publicly available soon on the Open Science Framework at https://osf.io/9vtws/. The setup featured a Meta Quest VR headset connected via a link cable to ensure stable performance. Moreover, participants used VR controllers to interact within the environment. Audio settings were standardized at 80% volume, and all notifications were muted to prevent distractions. Prior to each session, the VR system underwent boundary and ground level calibration to ensure consistency and safety across participants.

Procedure

Figure 1

Study Procedure Flowchart


Pre-Arrival Setup
35 minutes
Participant Arrival
1.Attaching wearable (EmbracePlus)
2.Walk up and down stairs five times
3.Register height, hip circumference, and skin tone
15 minutes
Baseline Measurements in-vivo
Standing Baseline (2 minutes)
Seated Baseline (2 minutes)
Lying Baseline (2 minutes)
Virtual Reality Stress Tasks
Randomization of
VR-TSST and Plank Walk
Seated Baseline
in VR (2 minutes)
Standing Baseline
in VR (2 minutes)
VR-TSST
Virtual job interview &
mental arithmetic
15 minutes
Plank Walk
Fear of heights scenario
10 minutes
Catching Butterflies
Cognitive control task
5 minutes
Self-reported presence
questions
End

Note: Flowchart illustrating the study procedure, including pre-arrival setup, participant preparation, sensor attachment, baseline measurements, three virtual reality tasks (VR-TSST, Plank Walk, and Catching Butterflies), and post-task assessments. After randomization the second VR task follows, e.g. if the plank task was randomized first the speech will follow and vice versa.

Pre-Arrival Setup procedure

Before participant arrival, all equipment was prepared including charging and connecting all devices, setting up the VR headset with boundary creation and ground level setting, and preparing the BIOPAC system with MP160, EDA receiver (A3), and AMI100D connections (BIOPAC Systems, Inc, 2024). The Unity application was launched, and the connection with AcqKnowledge was verified, which is a software platform developed by BIOPAC Systems that allows for the acquisition, visualization, and analysis of physiological data, such as ECG, EDA, EMG, and other bio signals. It is commonly used in research to record and process data collected from BIOPAC hardware devices.

Participant Arrival and Setup procedure

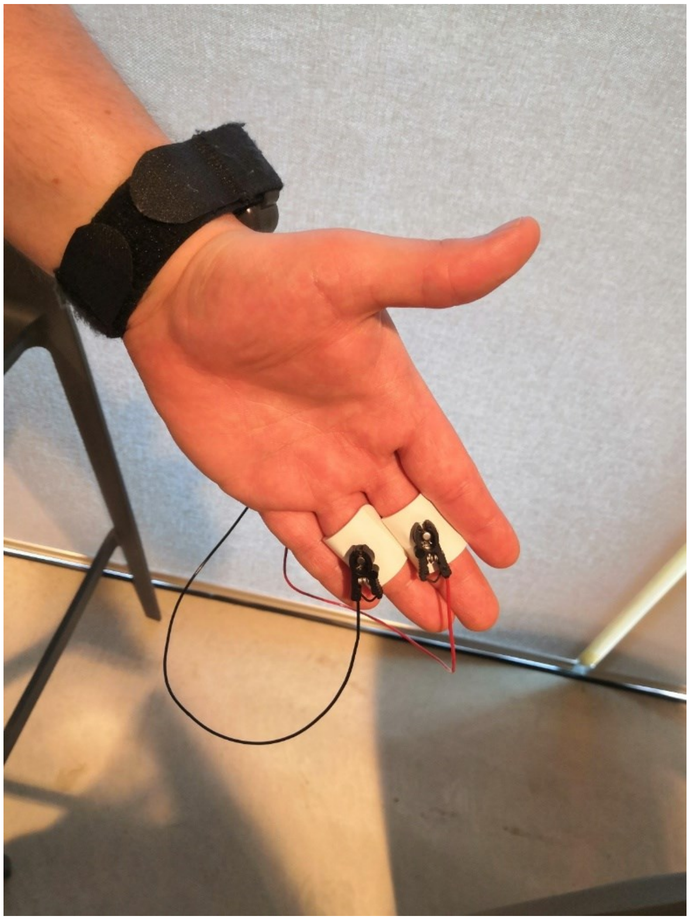
Upon arrival, the study was briefly explained, and participants were reminded of their right to withdraw at any time. Wearable devices were then fitted, followed by a standardized physical activity task requiring participants to walk up and down stairs five times (a total of 10 stair movements), in order build up some sweat for better conductivity of the devices. After physical activity the height in cm, weight in kgs, hip circumference in cm, and skin tone were registered.

Physiological Sensor Attachment procedure

The skin was gently abraded for ECG and ICG sensors, but not for EDA, and suitable conductive gels were applied to ensure optimal signal quality for ECG and EDA. Electrocardiogram (ECG) electrodes were placed on the right collarbone, the left lowest rib, and the right side of the belly button to enable accurate cardiac monitoring. Impedance cardiography (ICG) sensors were positioned on the sides of the neck and the ribs to measure changes in thoracic impedance. For electrodermal activity (EDA) measurement, sensors were applied to two fingers typically the middle and ring fingers or, using isotonic gel to ensure consistent conductivity (see Figure 2 below).

Figure 2

Electrodermal activity (EDA) sensors



Baseline Measurements procedure

Before participants were introduced to the virtual reality (VR) environment, a standardized pre-exposure procedure was followed. First, participants were asked to read a neutral text about wetland conservation, adapted from a Wikipedia article (Wikipedia Contributors, 2019), to ensure a neutral starting point. They then completed baseline questionnaires, including the International Physical Activity Questionnaire and the Pittsburgh Sleep Quality Index, to assess individual differences in physical activity and sleep quality. This was followed by a series of in-vivo baseline measurements. Participants began by standing quietly while baseline data were collected, after which they provided a subjective stress rating on a visual analogue scale (VAS), ranging from 0 (no stress) to 100 (maximum stress). Next, participants were instructed to sit quietly for two minutes without talking or moving to allow for baseline seated physiological measurements. This was followed by a lying down position for further baseline data collection. Once these measurements were complete, participants were fitted with the VR headset. Inside the VR environment, the baseline protocol continued with a seated baseline phase, followed by another VAS stress rating to assess subjective stress levels within the virtual setting. Finally, participants completed a standing baseline phase in VR to conclude the pre-task procedure.

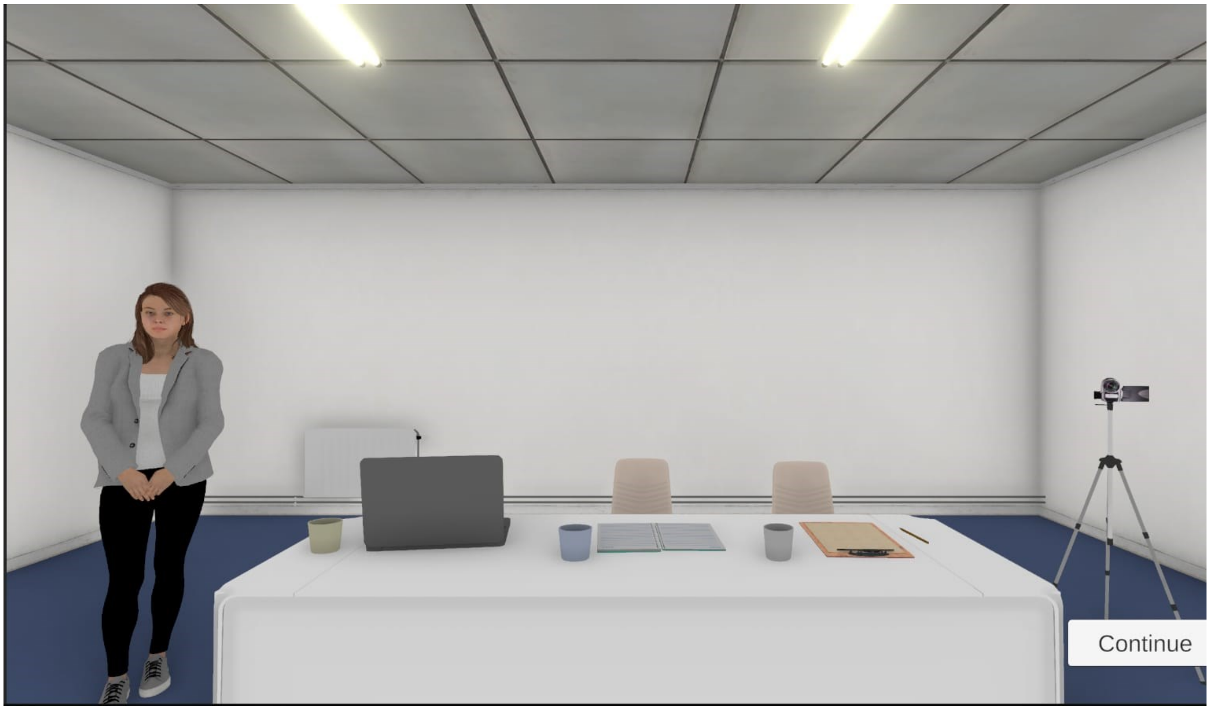
Virtual Reality Stress Tasks procedure

VR-TSST

The Virtual Reality Trier Social Stress Test (VR-TSST) closely followed the setup of the usual Trier Social Stress Test, but it was changed to work in a virtual world. The protocol consisted of three distinct phases. Before the Anticipation Phase, the participants received instructions from a virtual agent and were asked to prepare a five-minute speech about their strength and weaknesses for a simulated job interview (see Figure 3 below). This was followed by the five minutes Speech Phase, in which participants delivered their speech to virtual judges consisting of three evaluators sitting in front of the participant who maintained neutral or subtly disapproving facial expressions and body language. After the speech a Mental Arithmetic task followed for five minutes, in which they were tasked with performing serial subtraction, subtracting seven from 1,022 aloud, and were required to restart each time they made an error or if they were too slow, receiving immediate verbal feedback. This structure was designed to elicit a robust stress response in a controlled and standardized manner.

Figure 3

Virtual Reality Trier Social Stress Test (VR-TSST)



Note: Participants received instructions from a virtual agent and were asked to prepare a five-minute speech about their strength and weaknesses for a simulated job interview

Plank Walk Procedure

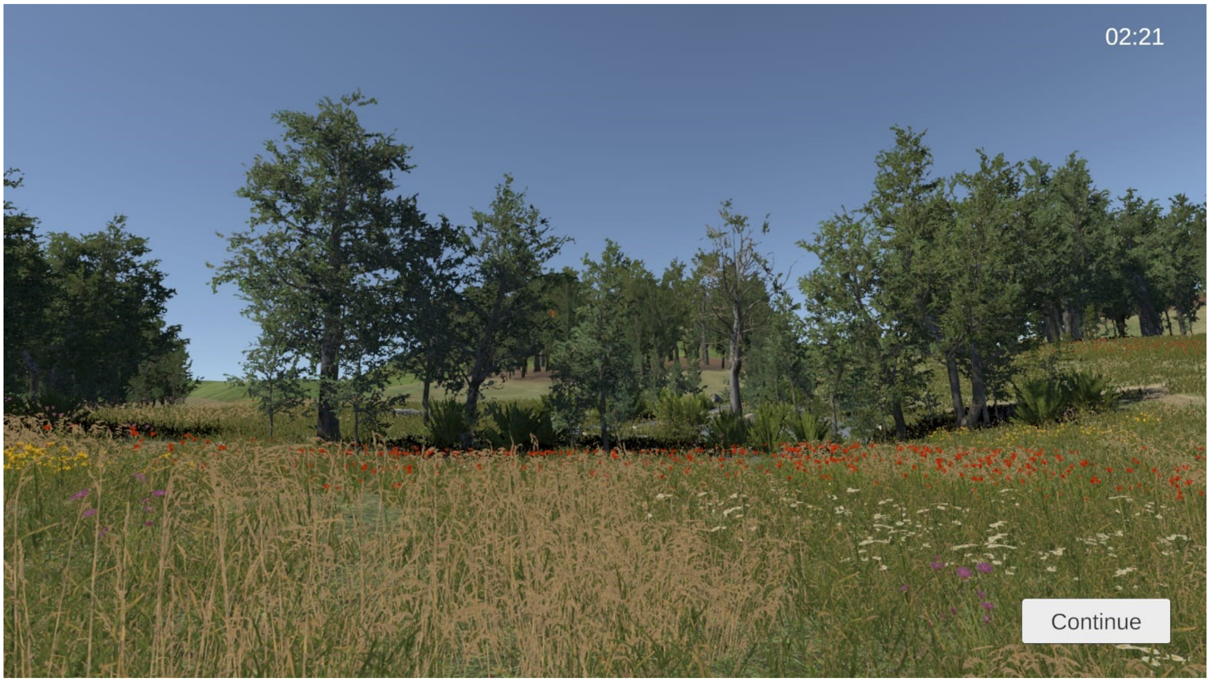
The Plank Walk was designed to induce a fear of heights in participants. At the beginning of the task, participants were placed in front of a virtual elevator and instructed to step inside. The elevator then ascended to a high floor equivalent of 80m high. Upon arrival, participants were instructed to walk at their own pace across a narrow wooden plank that extended over a virtual city. Once they reached the end of the plank, they were asked to stand still for two minutes and look down. To enhance immersion and elicit discomfort, the virtual environment featured realistic visuals of height and ambient wind sounds.

Catching Butterflies Procedure

This task served as both, a recovery and a movement inducing scenario in a nature-themed virtual environment. Participants were placed in a calm, open virtual space surrounded by trees and natural sounds, with the objective of touching butterflies that appeared randomly and varied in size, speed, and movement patterns (see Figure 4 below).

Figure 4

Catching Butterflies



Note: Nature-themed virtual environment, with the aim to replicate Netherlands' typical nature.

Post-Task Assessments procedure

After completing all virtual reality tasks, participants were asked to fill out a series of post-experiment questionnaires. These questionnaires assessed their subjective experience during the tasks, including their perceived sense of physical and social presence within the virtual environment. These scaled questionnaires can be reviewed in the appendix with the corresponding subscales included. Moreover, the questions have been systematically reviewed in the context of virtual reality evaluation, as highlighted in the systematic scoping review by Bareišytė et al. (2024) Additionally, participants reported their levels of fear related to heights and public speaking, as well as any feelings of motion sickness experienced during the tasks.

Data Analysis

The study utilized a modified version of the van Lier (2019) validation protocol to examine the similarity of electrodermal activity between the EmbracePlus smartwatch and the BIOPAC gold-standard system. The Correlation values were categorized using standard criteria: very weak (0.00–0.19), weak (0.20–0.39), moderate (0.40–0.59), strong (0.60–0.79), and very strong (≥ 0.80). Furthermore, bias in skin conductance level measurements was assessed using parameter-level agreement with a Bland–Altman plot. Here, a mean difference within ±1.6 µS was considered acceptable, based on van Lier's (2020) recommendations, which were derived from biologically plausible ranges of tonic EDA variation. Finaly, both fixed and proportional biases were assessed by examining the mean difference and the pattern of residuals, respectively.

To assess the relationship between subjective presence and changes in SCL, mean skin conductance levels were extracted for each participant during baseline and task conditions. This was done separately for two virtual reality tasks: the height-based plank task and the socially evaluated speech task. For the plank task, the baseline standing condition was used as the reference, as both the baseline and task involved a standing posture. For the speech task, the baseline seated baseline was used, reflecting the seated posture shared between baseline and task. For each task and participant, the change in average SCL was computed by subtracting the baseline value from the task value. This was done independently for both the reference device and the EmbracePlus. The presence scores were computed per participant and task. For the plank task, presence was measured using five items derived from the physical presence questionnaire (Bareišytė et al., 2024). For the speech task, the same five physical presence items were combined with an additional five items from the social presence questionnaire, reflecting the interpersonal nature of the speech task. Both questions with the corresponding subscales can be found in the appendix. These scores were then used as continuous predictors in subsequent regression analyses to investigate their relationship with SCL reactivity between both tasks. The data were compiled into long-format data frames for visualization and regression modelling. Separate linear regression models were conducted for each device (Biopac and EmbracePlus) and task to explore the association between presence and Δ SCL values.

Following, to examine whether subjective presence predicts variability in EDA signal agreement between the EmbracePlus and the reference device, a linear regression analysis was conducted using the cross-correlation values as the dependent variable. These coefficients represent the temporal alignment between the tonic EDA signals recorded by both devices across the entire duration of the virtual reality session. Like the speech phase, both the physical presence and social presence scales were combined to compute a composite presence score for each participant. This combined score served as the predictor in the regression model. Following, a simple linear regression was performed to evaluate the extent to which presence scores accounted for variability in signal agreement. Standard regression metrics such as slope, intercept, R^2, and p-value were calculated to assess the predictive relationship. All data processing, statistical analyses, and visualizations were carried out in Thonny 4.1.7 with Python version 3.10.11 and Windows 11. Finaly, a regression plot was created to illustrate the association between composite presence scores and EDA cross-correlation values.

Results

Electrodermal Activity Measurements

Substantial differences were observed between the EmbracePlus smartwatch and the BIOPAC gold-standard system across all electrodermal activity parameters. As shown in *Table 2*, the EmbracePlus produced much lower average skin conductance level (SCL) than the BIOPAC system, with the BIOPAC recording approximately 8.5 times higher mean SCL values. Notably, both devices showed considerable variability between participants, though the BIOPAC demonstrated a broader absolute range of responses. Further details are provided in Table 3, which depicts means and standard deviations of skin conductance levels (SCL) for both devices across standing and seated baselines, their corresponding stress tasks (plank and speech), as well as in-vivo baselines. The data indicate that mean SCL increased from baseline to the respective stress task for both systems. However, across all conditions, the absolute SCL values recorded by the BIOPAC remained substantially higher than those recorded by the EmbracePlus.

Table 2

Skin Conductance Level (SCL) across whole experiment

Measure	**EmbracePlus**	**Biopac**
SCL, M (SD), µS	1.22 (1.69)	10.45 (6.67)

Measure	EmbracePlus	Biopac
SCL Range, µS	0.01–8.60	1.47–37.31

Note. SCL = skin conductance level; M = mean; SD = standard deviation; µS = microsiemens. Values reflect participant-level averages across all phases.

Table 3

Skin Conductance Level (SCL) for Standing/Seated Baselines and Corresponding Stress Tasks (Mean ± SD, in µS), with Change Scores (ΔSCL)

Condition Pair	Biopac SCL (µS)	EmbracePlus SCl (µS)	ΔSCL (Biopac)	ΔSCL (EmbracePlus)
Standing Baseline in VR	8.47 ± 4.68	0.99 ± 1.26		
Standing Baseline in-vivo	9.69 ± 4.60	1.18 ± 1.62		
Plank Task	9.95 ± 4.91	1.07 ± 1.31	1.49	0.07
Seated Baseline in VR	8.59 ± 4.31	0.94 ± 1.19		
Seated Baseline in-vivo	8.53 ± 4.37	1.40 ± 1.57		
Speech Task	9.58 ± 4.85	1.20 ± 1.58	0.99	0.26

Note. Δ = Plank task minus standing baseline in VR. Δ = Speech task minus seated baseline in VR. Means and SDs are based on participant-level averages across devices and conditions.

Presence Questionnaires

Participants generally reported moderate levels of both social and physical presence in the virtual environment. The Social Presence scale (5 items; α = .78) produced a mean score of 3.20 (SD = 0.68) on a 5-point Likert scale, indicating a general sense of co-presence with the virtual characters. The highest-rated item was “I felt that the people in the virtual environment were aware of my presence” (M = 3.79, SD = 0.66), suggesting that participants felt somewhat acknowledged by the virtual others. In contrast, the lowest-rated item was “The people in the virtual environment appeared to be sentient” (M = 2.15, SD = 0.81), reflecting a limited sense of the virtual agents being truly lifelike or conscious.

On the Physical Presence scale (5 items; α = .81), participants reported a similar level of immersion, with a mean score of 3.11 (SD = 0.73). The item most strongly endorsed was “While I was in the virtual environment, I had a sense of being there” (M = 3.64, SD = 0.69), indicating that the environment successfully evoked a basic sense of spatial presence. However, the statement “I was completely captivated by the virtual world” received the lowest rating (M = 2.47, SD = 1.02), suggesting that full absorption into the experience was less common. Taken together, the results show that while the VR setup was effective in generating a moderate subjective sense of presence, it fell short in creating a fully immersive or socially convincing experience.

Post-experiment questionnaire

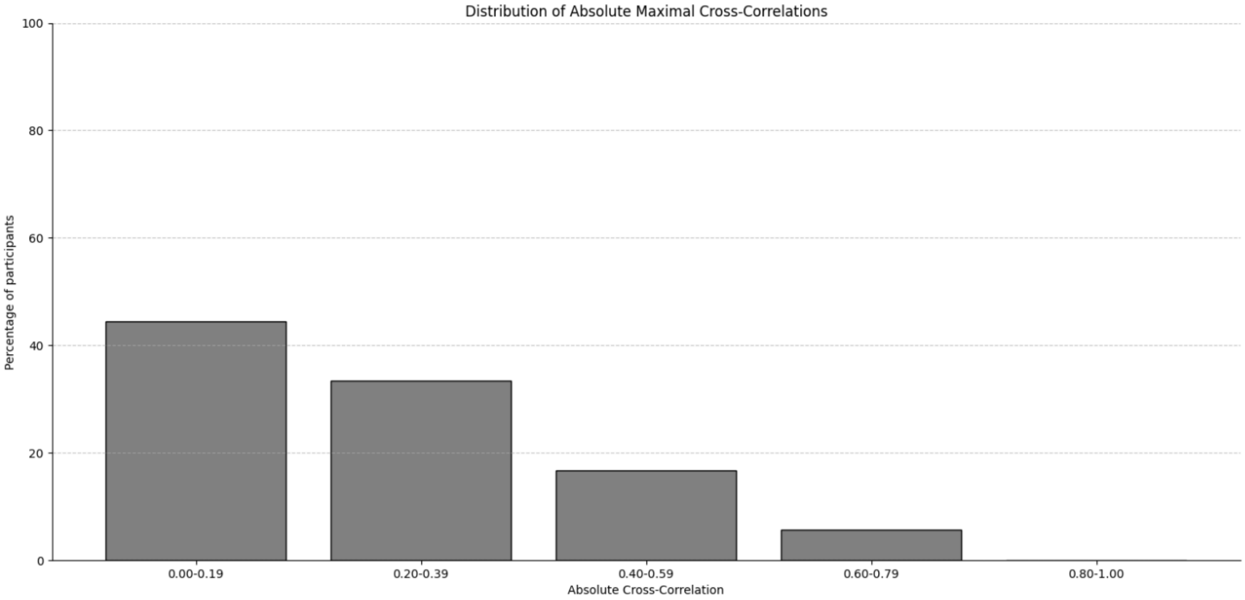
In addition to the standardized presence questionnaire, participants were asked several open-ended questions about their personal experiences during the VR sessions. While the responses provided valuable context and helped illustrate how participants experienced the tasks, they were not directly relevant to the main research questions and were therefore not included in the core analysis. Still, for readers interested in these reflections such as how participants described stress during the speech task, their reactions to the nature scenario, a full overview of the synthesized responses can be found in the Appendix: Post-Experiment Questionnaire.

Inferential statistics

The first research question examines the measurement validity of the EmbracePlus smartwatch in recording EDA during the VR tasks. It investigates how closely the smartwatch's EDA measures, particularly skin conductance level corresponds to those recorded by the reference device. Results showed low overall agreement, with a mean correlation of $r = .19$ (SD = .29, range = –.53 to .80). Notably, only one participant met the validity threshold of $r > .80$. As shown in Figure 5, 41.2% of participants fell into the "very weak" category (r = 0.00–0.19), 35.3% showed "weak" correlations (r = 0.20–0.39), 17.6% "moderate" (r = 0.40–0.59), and only one participant (5.9%) reached the "strong" range (r = 0.60–0.79). No systematic time lag was observed across participants. Figures 6 and 7 illustrate the best (Participant 07: $r = .797$, lag = 0.19s) and worst (Participant 02: $r = –.53$) alignment cases, respectively. The optimal lag varied between participants, ranging from -8.00 to 8.00 seconds, and included both positive and negative values. The average lag across participants was 0.078 seconds (SD = 5.99 seconds), indicating high variability in alignment. All cross correlations can be found in (Appendix1).

Figure 5

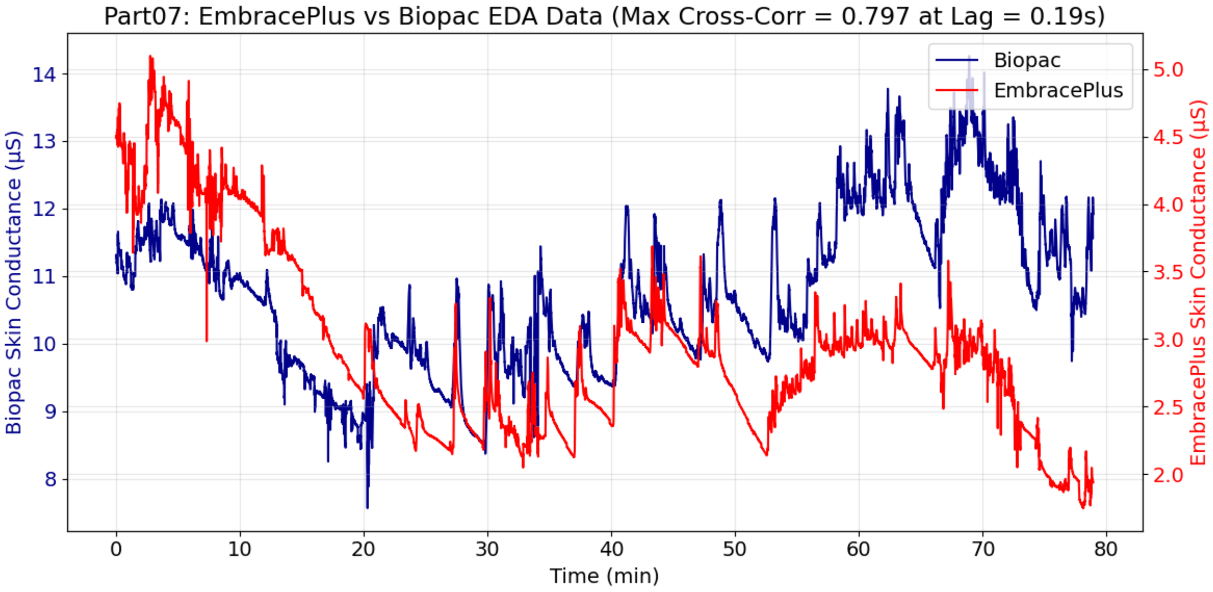
Distribution of Absolute Maximal Cross-Correlations



Note: Distribution of absolute maximal cross-correlations between the EmbracePlus smartwatch and BIOPAC gold-standard electrodermal activity (EDA) measurements across all participants and virtual reality (VR) stress tasks. Values represent the percentage of participants whose maximal cross-correlation fell within each absolute cross-correlation range.

Figure 6

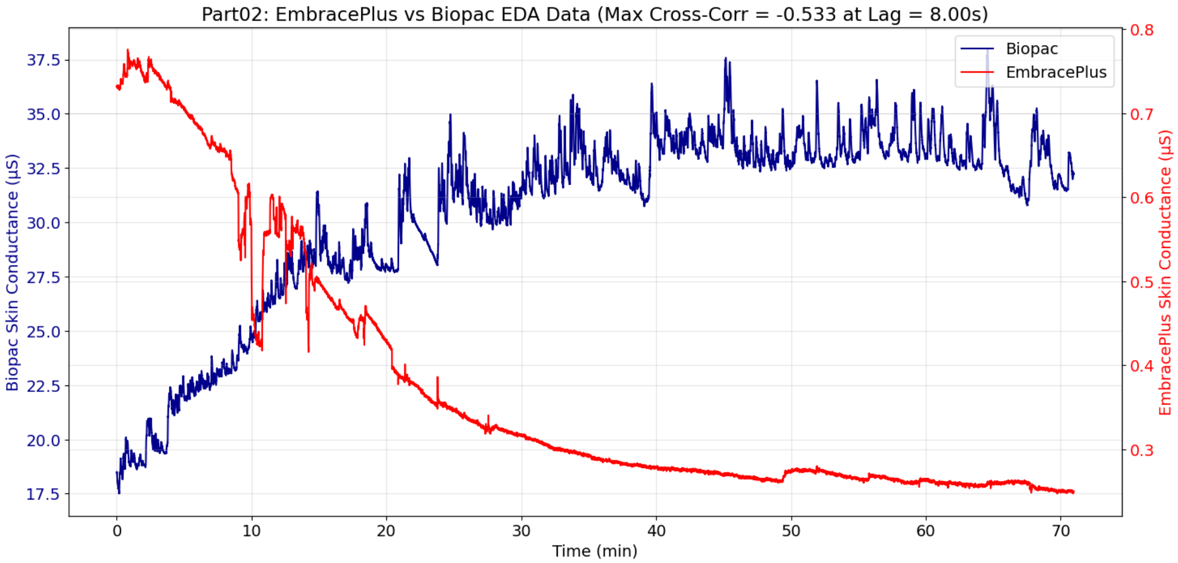
Participant 7 (Cross correlation = 0.797)



Note: This figure displays the skin conductance data from Participant 7, comparing recordings from a Biopac system (RD Biopac) and the EmbracePlus. A maximum cross-correlation of 0.797 was observed between the two signals at a lag of 0.19 seconds. The left y-axis represents the values for the BIOPAC, and the right y-axis illustrates the values for the EmbracePlus.

Figure 7

Participant 2 (Cross correlation = -0.533)



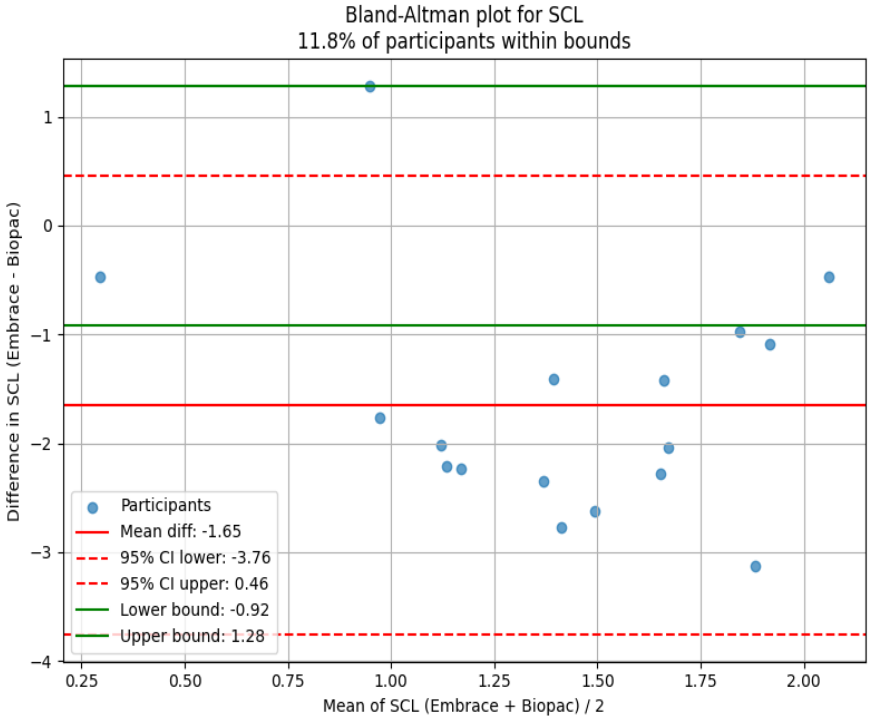
Note: This figure displays the skin conductance data from Participant 2, comparing recordings from a Biopac system (RD Biopac) and the (EmbracePlus. A maximum cross-correlation of -0.533 was observed between the two signals at a lag of 8.00 seconds. The left y-axis represents the values for the BIOPAC and the right y-axis illustrates the values for the EmbracePlus.

Parameter Level

To evaluate whether the EmbracePlus systematically differs from the Biopac system in the measurement of SCL, a Bland–Altman analysis was conducted. The predefined criterion for acceptable agreement was set at ±1.6 µS. The analysis yielded a mean difference of –1.65 between the two devices (SD = 1.08), with 95% confidence intervals ranging from –3.76 to 0.46. Only 11.8% of data points fell within the acceptable limits of agreement (–0.92 to 1.28 in log-transformed units). Lastly, the results were back transformed into microsiemens (µS) for simpler interpretation. With a 95% confidence interval between -0.98 and 0.59 µS, the back-transformed mean difference was -0.81 µS. The corresponding Bland–Altman plot is shown in Figure 8 below.

Figure 8

Bland-Altman plot for SCL



Note: Bland-Altman plots illustrate log-transformed EDA parameters, comparing EmbracePlus and Biopac. SCL (11.8% within boundaries. The plot displays mean differences, 95% confidence intervals, and predefined acceptable boundaries.

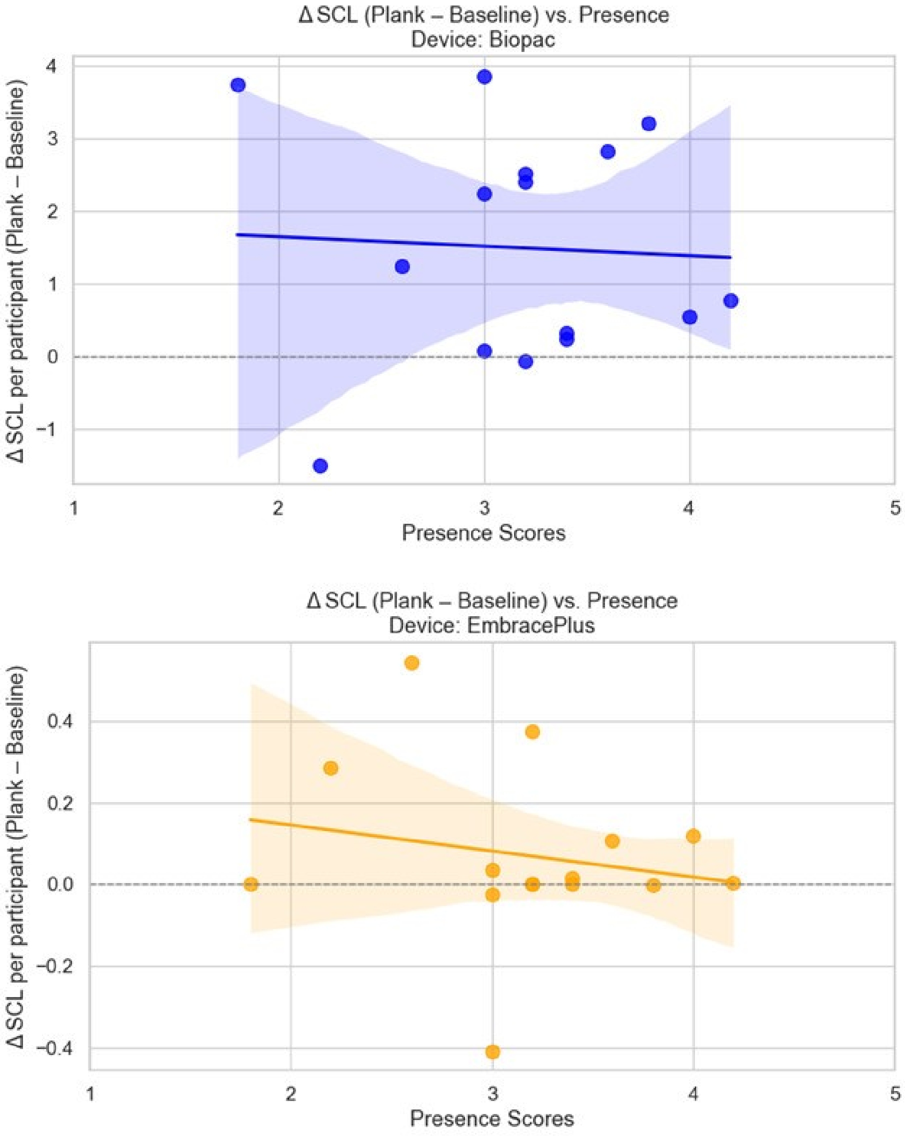
Presence effect

To test whether self-reported presence predicted changes in skin conductance level (ΔSCL) across VR phases, ΔSCL values were computed for each participant and device (Biopac, EmbracePlus). Two comparisons were examined: (1) standing baseline vs. plank, and (2) seated baseline vs. speech. On average, ΔSCL was positive in both comparisons, indicating higher SCL during the stress tasks. For the plank task, mean ΔSCL was larger with the Biopac (M = 1.90 µS, SD = 1.51) than with the EmbracePlus (M = 0.27 µS, SD = 0.20). Regression analyses, however, showed no significant relationship between presence and ΔSCL. For the plank task: Biopac slope

= –0.13, $R^2 = .003$, $p = .85$; EmbracePlus slope = –0.06, $R^2 = .04$, $p = .50$. For the speech task: Biopac slope = 0.33 (SE = 0.52), intercept = –0.004, $R^2 = .03$, $p = .54$; EmbracePlus slope = –0.15 (SE = 0.26), intercept = 0.73, $R^2 = .03$, $p = .57$. Scatterplots in Figure 9 and 10 confirmed these null findings, showing no clear association between presence and ΔSCL. Overall, presence did not systematically influence SCL changes across conditions.

Figure 9

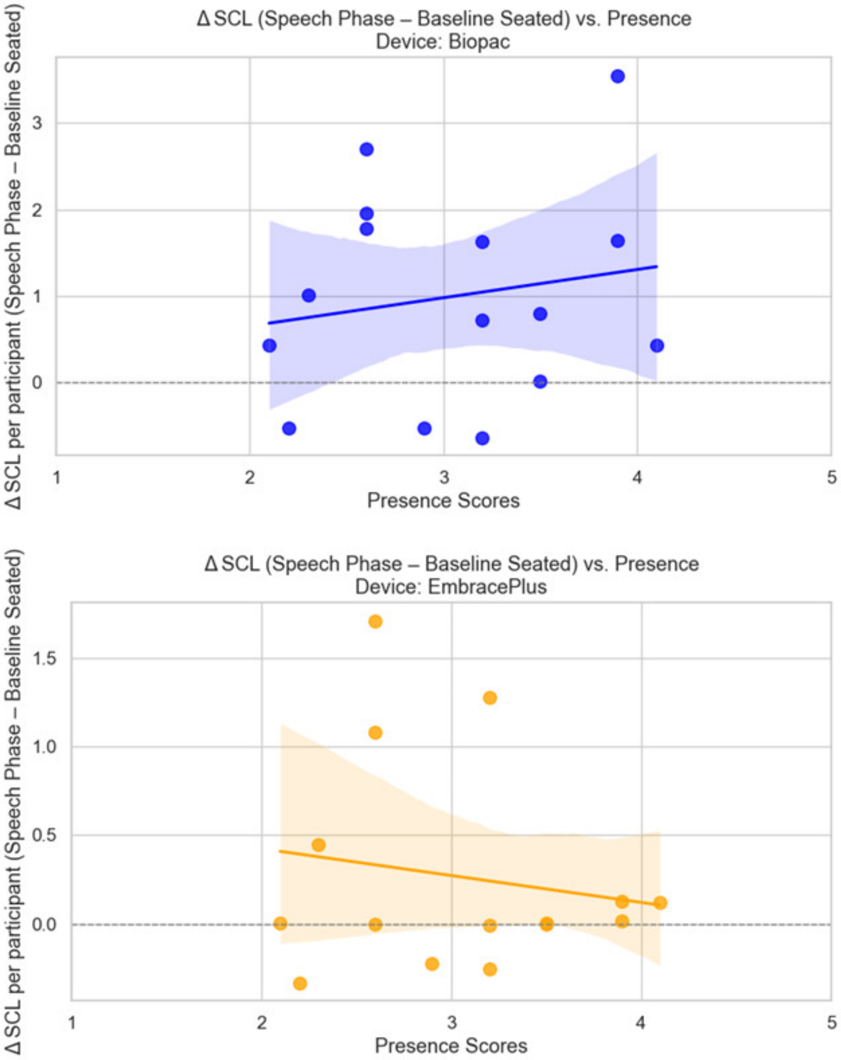
Relationship between presence and ΔSCL during the plank task (Biopac and EmbracePlus).



Note. The scatter plots show individual differences in skin conductance level (SCL) between the plank task and the standing baseline as a function of presence scores, recorded with both Biopac (blue line) and EmbracePlus (orange line). Each dot represents a participant, and the regression lines indicate linear fits with 95% confidence intervals. Higher ΔSCL values reflect stronger physiological arousal during the plank task. Presence scores ranged from 1 (low) to 5 (high).

Figure 10

Relationship between presence and ΔSCL during the speech task (Biopac and EmbracePlus).



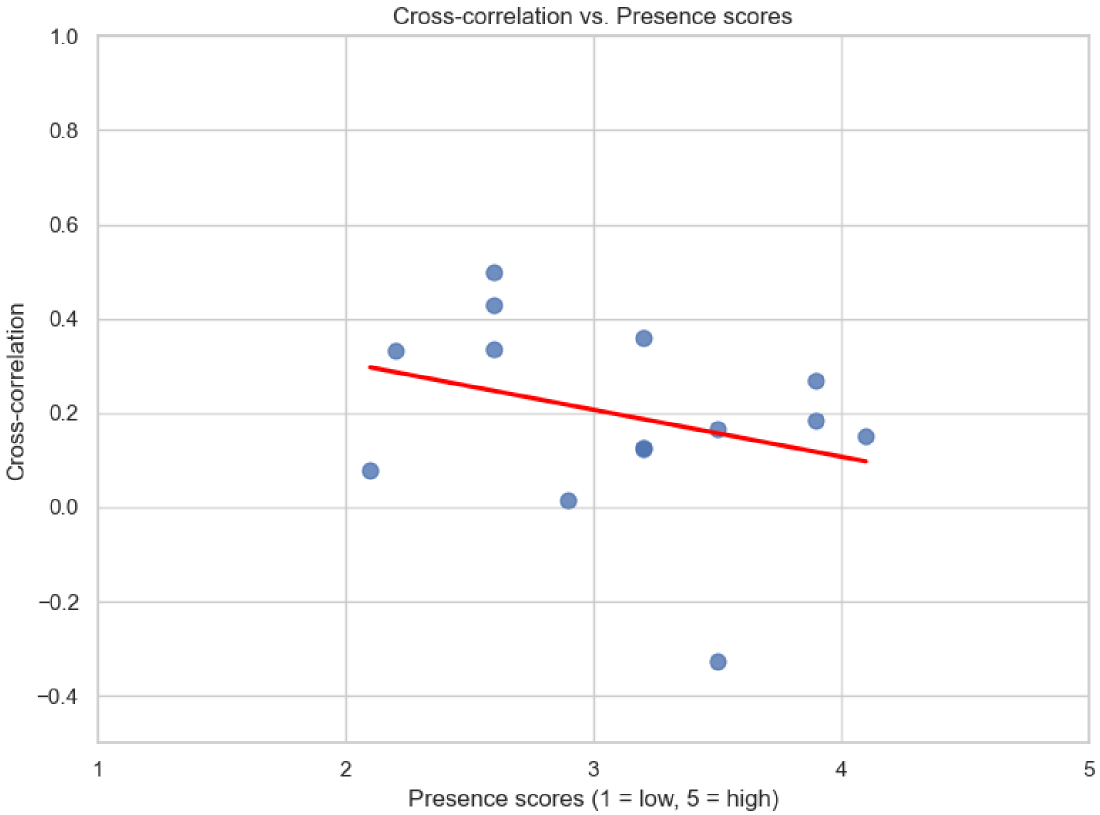
Note. The scatter plots show individual differences in skin conductance level (SCL) between the speech task and the seated baseline as a function of presence scores, recorded with both Biopac (blue line) and EmbracePlus (orange line). Each dot represents a participant, and the regression lines indicate linear fits with 95% confidence intervals.

Moderating Role: Relationship Between Presence and EDA Agreement

One participant (P07) exhibited an unusually high cross-correlation (r = 0.80), well above the group average and exerted strong influence on the regression model. To avoid distortion in this small sample, the primary results are reported with this participant excluded. Without P07, the regression analysis indicated no significant relationship between self-reported presence and EDA signal agreement between the EmbracePlus and BIOPAC (slope = –0.10, $R^2 = 0.09$, p = .28). Visual inspection (Figure 11) did not reveal any systematic association. For transparency, including P07 in the analysis resulted in a somewhat stronger model fit ($R^2 = 0.18$, slope = –0.17, p = .11), but this apparent effect was driven primarily by that single data point.

Figure 11

Cross-correlation vs. presence scores



Note: Each data point corresponds to one participant. Participant 7 was identified as an outlier, with an unusually high cross-correlation value (> 0.7), and was excluded from the analysis to avoid skewing the results. With this participant included, the relationship was not statistically significant (p = 0.11).

Discussion

Main Findings

The purpose of this study was to gain a better understanding of the validity of the EmbracePlus smartwatch for measuring electrodermal activity (EDA) during virtual reality stress tasks and to examine the potential moderating role of presence in device validation. There are three key findings that address the core research questions.

To begin with, when looking at the signal level, the results showed that the agreement between the two devices was generally low. Most participants had only weak or very weak correlation between the smartwatch and the reference device, and only one participant showed a strong correlation. Furthermore, some participants even showed negative relationships, indicating that the EmbracePlus was not capturing the same signal patterns. In addition, in the parameter-level the Bland–Altman plot revealed that the EmbracePlus consistently recording lower SCL in comparison to the BIOPAC system. To, specify, only a small number of measurements fell within the acceptable range of agreement. This same pattern is visible in the cross-correlations and suggests that while the smartwatch might be able to measure EDA, it tends to underestimate the intensity of skin conductance responses. However, this was expected since the EmbracePlus is using dry electrodes and is placed on the wrist instead of the finger like the reference device.

Following, the study explored whether participants' sense of presence was related to changes in skin conductance levels (SCL) across two different tasks, the *plank* and *speech* phases. Even though the statistical analyses did not show a significant relationship, the scatterplots from both devises showed that most data points were in the positive range. This means that the SCL was generally higher during the plank and speech phases compared to the standing and seated baseline, which was expected given the stress-inducing nature of the tasks. Hence, the VR tasks successfully created a convincing and effective increase in EDA, which is a good finding for the long-term use of the VR environments. Comparing the two tasks with each other, the mean SCL values of the EmbracePlus mean SCLs were much smaller compared to the reference device but still mostly above zero. The slope of the relationship between presence and SCL change was slightly negative for the plank phase, and a slightly positive for the speech phase measured by the Biopac. However, this trend should be interpreted with caution due to the small sample size, but it is still interesting to observe because it indicates that participants who felt more present in the virtual environment showed slightly less of a stress response. This could mean that strong stress reactions can still occur even when the persons sense of presence is relatively low, which is useful to consider when designing future VR stress protocols.

Lastly, the study examined whether presence affected the agreement between the smartwatch and RD´s EDA measurements regarding cross-correlation. Notably, presence did not seem to make much of a difference when looking at the plot, but also here it important to mention that there was no big sample size and the effect was not significant. The linear regression analysis revealed that participants with higher presence scores did not show better agreement between the two devices. In fact, there was a small negative trend, meaning that those who felt more present had slightly lower correlations between the devices. Hence, it challenges the idea that feeling more present in the virtual environment automatically leads to more accurate or reliable physiological recordings from wearables. In conclusion, the presence did not seem to moderate the alignment between the EmbracePlus and the reference system in this study. However, it was clear that the VR environment could produce visible phase changes for most participants, regardless of presence score, the VR experiment managed to generate EDA arousal from the key activities plank and speech.

Comparison to prior work

One finding of the signal level was the low agreement between the EmbracePlus smartwatch and the RD when comparing EDA signals. Most participants showed weak or very weak correlations between the two devices (average $r = .19$), with only one participant demonstrating a strong correlation. This result is consistent with earlier findings by van Lier et al. (2020), who also reported similar fining with limited agreement between the older E4 device and the reference device. Potential factors for the low correlations can include differences in sample size, sensor placement, and participant movement during the VR tasks, which can change physiological signals (McDuff et al., 2024). Unlike reference devices that use wet electrodes placed on the fingers or palms, which provide high sensitivity to subtle conductance changes, the EmbracePlus relies on dry electrodes positioned at the wrist, where fewer sweat glands are located (Boucsein, 2012; van Dooren et al., 2012). This setup is more susceptible to noise and motion artefacts, particularly during physically engaging tasks (McDuff et al., 2024). In addition, sampling frequency may also play a role, as lower sampling rates reduce the device's ability to capture quick changes in electrodermal activity, further decreasing alignment with reference measures (Menghini et al., 2019). Taken together, these technical limitations provide a plausible explanation for why the EmbracePlus, despite being a newer generation wearable, still struggles to achieve strong agreement with laboratory-grade systems. These findings also fit with wider challenges reported in the Stress in Action Wearables Database (Schoenmakers et al., 2025). The database highlights considerable variation in reported validation statistics such as Bland–Altman plots, intra-class correlations, and regression metrics, which can complicate cross-study and cross-device comparisons. Furthermore, inconsistent reporting and limited manufacturer transparency regarding signal processing and sampling specifications were also noted as major problem in evaluating wearable data quality (Canali et al., 2022). What´s problematic is, that many physiological parameters reported by smartwatches lack proper validation, and reference devices mostly outperform them in terms of reliability and validity. Similarly, Peake et al. (2018) pointed out that many commercial wearable devices still lack proper validation. Even though the EmbracePlus is a newer model, it seems to have similar limitations. This highlights the need for continued standardized testing, especially as smartwatch technology continues to advance quickly.

Regarding the parameter-level analysis using Bland–Altman plots showed that the EmbracePlus consistently recorded lower SCL than the BIOPAC system. Only a small number of participants fell within the expected range of agreement. This result supports earlier findings that wearables often underestimate physiological measurements compared to reference devices (Kleckner et al., 2021). According to van Lier et al. (2020), parameter-level analysis helps to assess whether two devices measure the same quantity in a similar way. The findings suggest that while the smartwatch detected general increases in skin conductance during stress tasks, it did not match the size and range of changes recorded by the RD. This is important when considering the use of wearables in natural settings e.g. daily activities, where precise measurement is needed.

The third part of the study looked at whether the sense of presence in VR influenced changes in SCL between baseline and stress conditions. As prior expected, the SCL increased during the speech and plank tasks compared to the baselines, which fits with the idea that these tasks caused stress. However, no significant association was found between self-reported presence scores (both physical and social) and the magnitude of skin conductance level changes across VR tasks. This result differs from some earlier research suggesting that higher presence in VR can lead to stronger bodily responses (Grassini & Laumann, 2020; Lemmens et al., 2021). One possible reason for the difference in findings is that the present study did not actively manipulate presence levels. Instead, it simply measured how present participants felt while experiencing the same virtual environment. In contrast, earlier studies that did find an effect compared to very different types of environments such as virtual reality versus 2D images which likely created a much stronger contrast in participants' experiences (Siavash Eftekharifar et al., 2021). This suggests that the type of task or environment may influence stress responses more than the subjective feeling of presence alone.

Strengths & Limitations

This study offers several valuable contributions to the field of device validation, yet it also comes with strong and weak points that need to be acknowledged. To begin with, a strict validation procedure was used, adhering to the framework developed by van Lier (2020). This included multiple layers of analysis: signal-level with cross-correlation, and parameter-level with the Bland–Altman comparison, and further tests for presence as a moderating factor were added. By combining these analytical approaches, the study offers a multifaceted evaluation of device performance. The depth of the validation results was further enhanced by integrating not only objective data from the devices but also subjective reports from participants about their sense of presence. This combination provides a more comprehensive understanding of how device agreement relates to both physiological and experiential factors. Furthermore, using virtual reality for validation also marks an important step forward in how this kind of research is done. Unlike traditional lab-based studies that often employ simple stress tasks in bodily resting states, this research used VR environments that simulate real-world stressors while keeping experimental control (Canali et al., 2022). Lastly, exploring presence as a potential moderator in device validation offers an innovative contribution to the field. Although prior research has shown that presence affects physiological responses (Siavash Eftekharifar et al., 2021). This study is among the first to investigate whether psychological presence influences measurement validity between wearable and laboratory-grade systems.

Despite its strengths, this study also has some limitations that need to be considered. The study's sample size of 20 participants poses a limitation, impacting both the generalizability and statistical power of the results. Research on sample sizes for validation studies highlights that small sample sizes are especially problematic, as they can produce imprecise estimates of device performance and increase vulnerability to outliers (Dhiman et al., 2023). This issue is further complicated by the fact that only one participant exceeded the > 0.79 cross-correlation threshold, making it difficult to properly evaluate the device's performance under conditions that meet validation standards. Furthermore, the participant sample consisted only of university students aged 20 to 29, which limits the external validity of the findings for broader populations. Wearable device performance can differ across demographic groups due to variations in skin conductance, body composition, and age-related physiological factors (Hongn et al., 2025; Huang et al., 2022). This demographic homogeneity restricts the applicability of the results to older adults, individuals with varying health conditions, or those from diverse socioeconomic backgrounds. Moreover, the single-session cross-sectional design limits the ability to evaluate device performance over time or in varying contexts (Kooiman et al., 2015). Factors such as wear time, skin adaptation, and changing environmental conditions can affect wearable device accuracy across different situations. In addition, although three distinct VR tasks (VR-TSST, Plank Walk, Catching Butterflies) were used, the validation remained limited to specific stress-induction scenarios. To specify, real-world EDA responses cover a far wider range of physiological states and contexts, such as everyday activities, sleep, and diverse emotional experiences (Boucsein, 2012). Therefore, the task-specific focus restricts the ability to generalize conclusions about device performance across the full range of time user´s wearing the watch. Lastly, the validation focused on electrodermal activity, which limits the ability to draw conclusions about the EmbracePlus device’s performance in measuring other physiological parameters e.g. heart rate or heart rate variability.

Implications for Future Research and Practice

This validation study provides a base for future work aimed at improving smartwatch and VR technology. A key consideration for follow-up research is the limited variability in presence scores observed in the current sample, which was skewed toward high presence. This may have constrained the ability to detect a specific relationship between presence and SCL. Future studies could aim for greater variability in presence experiences for instance, by using tasks with varying levels of immersion or by recruiting participants with differing familiarity with VR. Moreover, given the moderate sample size, replication with larger and more diverse populations could be beneficial to confirm these findings. A future student building on this work might explore whether participants with very low presence should be excluded or analysed separately. From a practical perspective, the consistent underestimation of electrodermal activity by the EmbracePlus device presses the need for improved signal processing algorithms. These could better account for individual physiological differences and environmental influences. Further, continued collaboration between researchers and device manufacturers could help enhance wearable accuracy, particularly in dynamic, real-world settings such as VR.

Conclusion

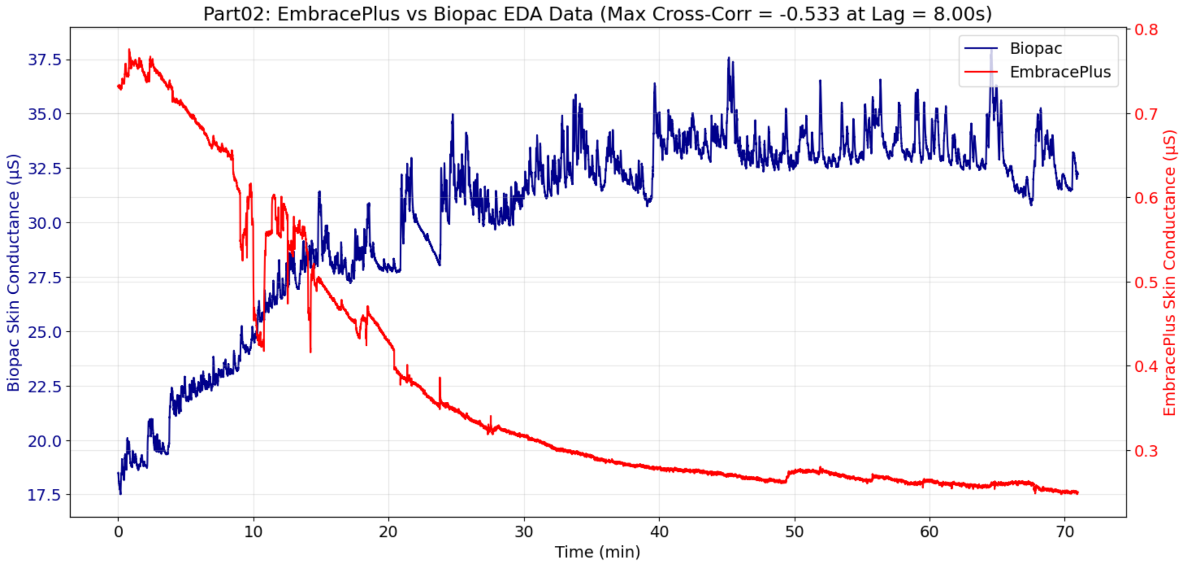
This study offered a detailed assessment of the EmbracePlus smartwatch’s ability to measure electrodermal activity (EDA) during VR stress tasks, while also identifying key limitations in its current performance. Compared to the gold-standard BIOPAC system, the EmbracePlus showed weak agreement at both the signal and parameter levels, consistently underestimating skin conductance and demonstrating poor temporal alignment. Additionally, the expected moderating effect of presence was not observed, indicating that the subjective sense of presence may not significantly influence the accuracy or agreement of wearable EDA data in immersive VR settings. Despite some limitations, this study adds to the field by using a strong, multi-level validation approach in realistic settings. By using VR, it also explores presence as a key factor influencing stress responses something past studies without VR haven’t looked at. That said, the small sample and cross-sectional design for presence limit how broadly the results can be applied. Future research should include more diverse participants and track changes over time. Overall, the results show the ongoing challenge of getting research-level accuracy from consumer wearables, and point to the need for better sensors, smarter data processing, and consistent validation methods suited for real-world use. Importantly, this study successfully applied a validation protocol developed by the broader *Stress in Action* project team, providing preliminary evidence that immersive environments can be used to systematically evaluate wearable stress sensors under ecologically valid conditions. The consistently moderate-to-high levels of presence reported by participants confirm that the VR setup was effective in creating immersive stress-inducing scenarios, supporting its continued use in future validation studies.

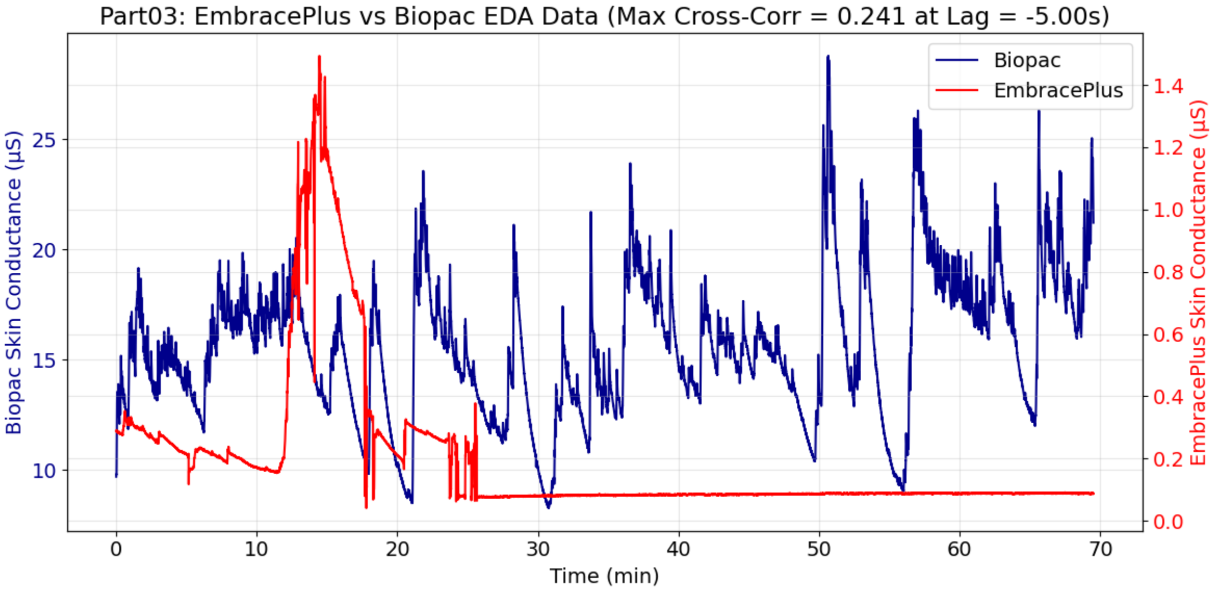
Appendix

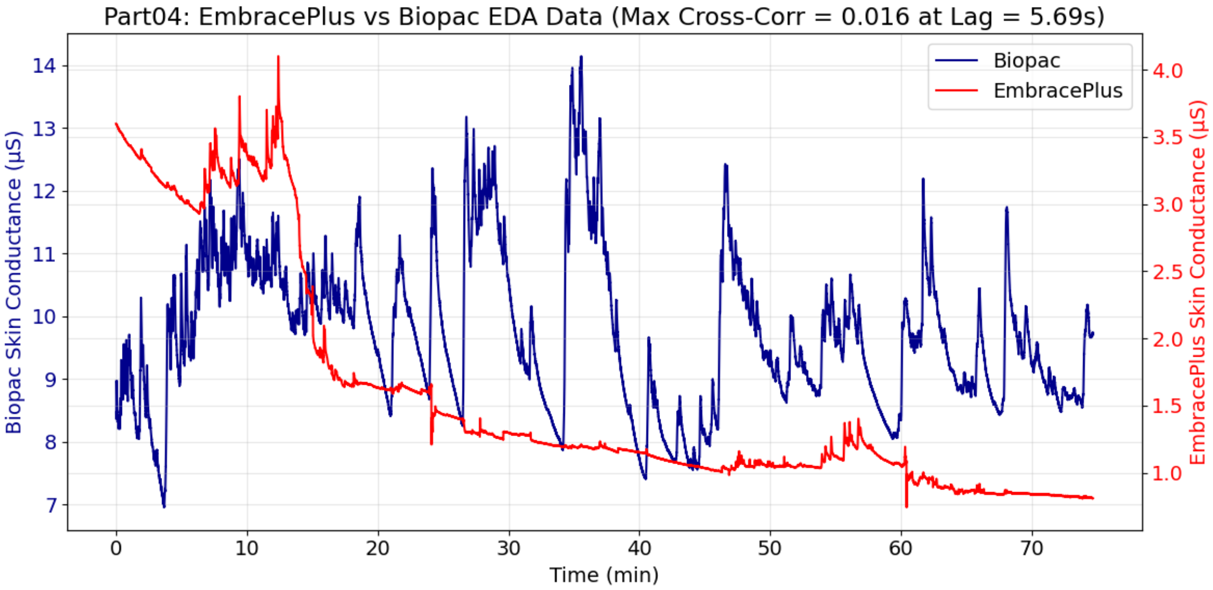
Consent form:

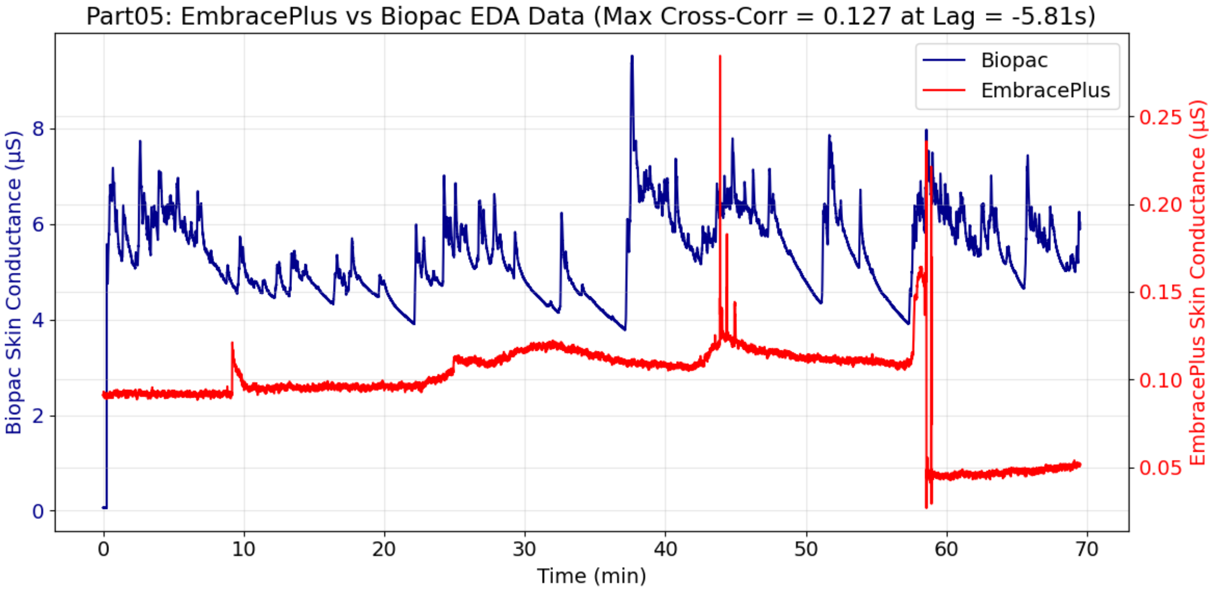
https://utwentebs.eu.qualtrics.com/jfe/form/SV_5Bj03AnlXqoy18i

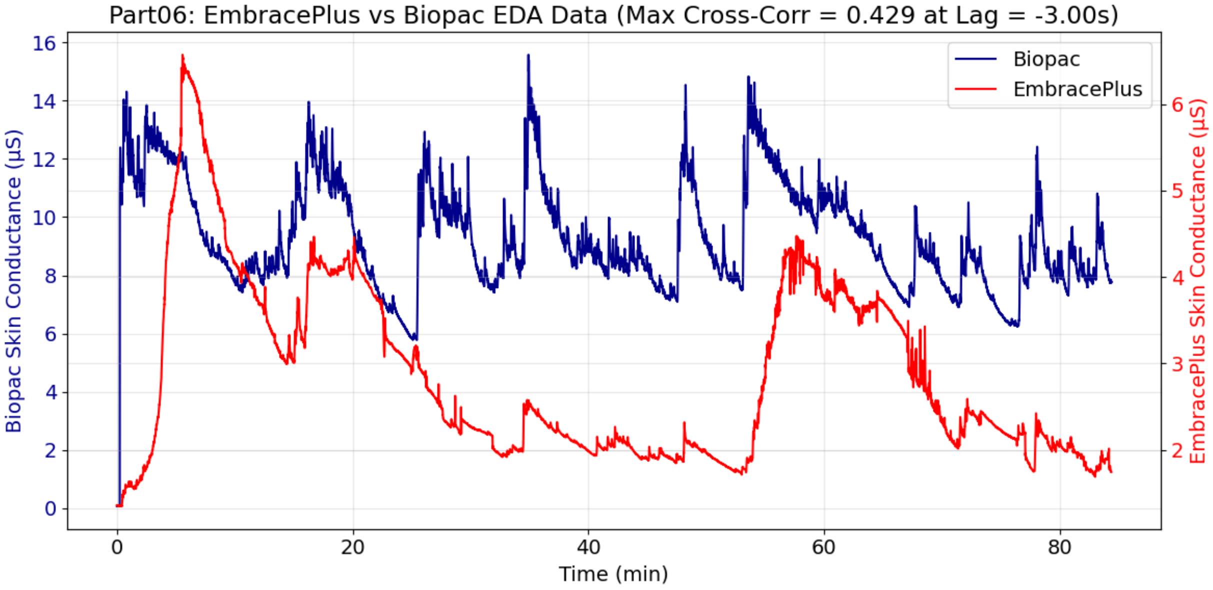
All Cross-correlations:

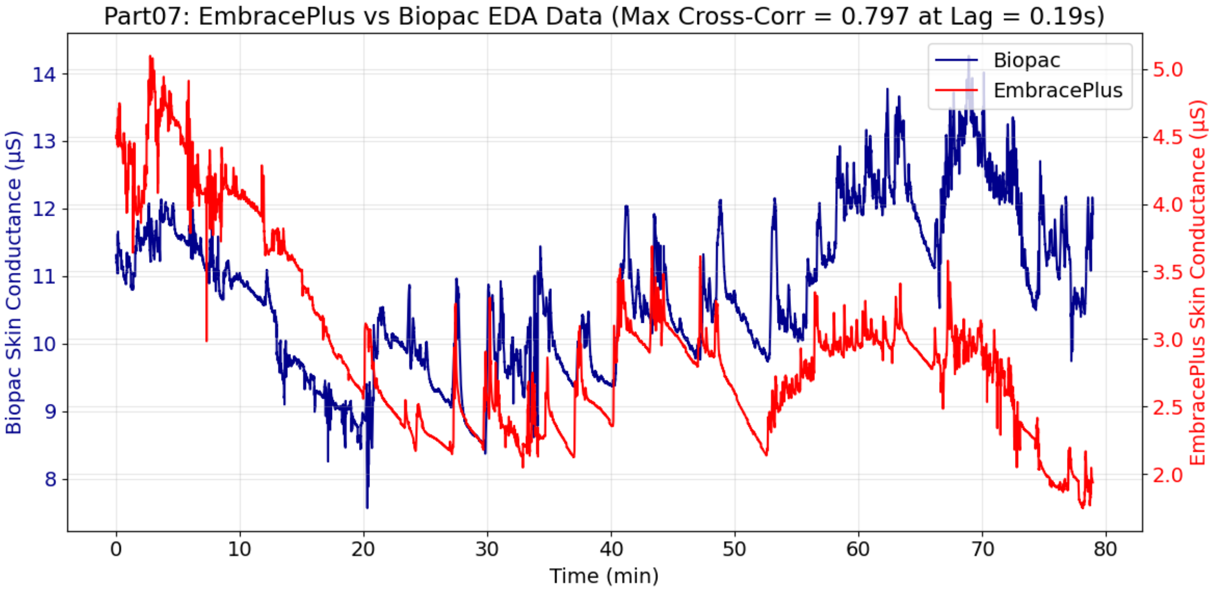


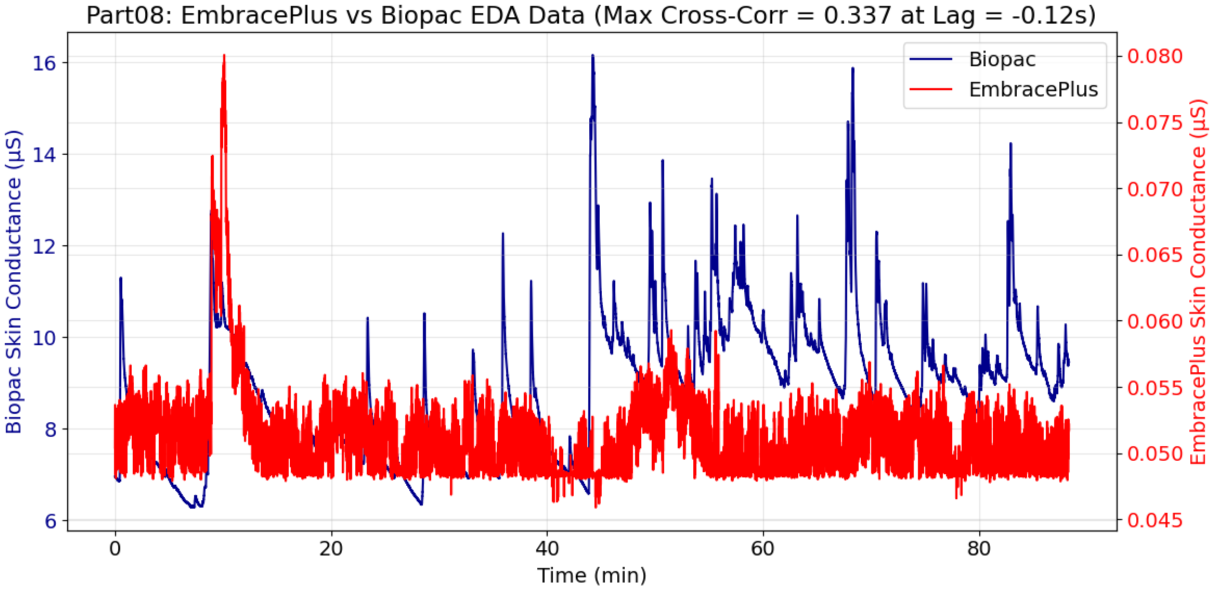

Part03: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.241 at Lag = -5.00s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
10
20
30
40
50
60
70

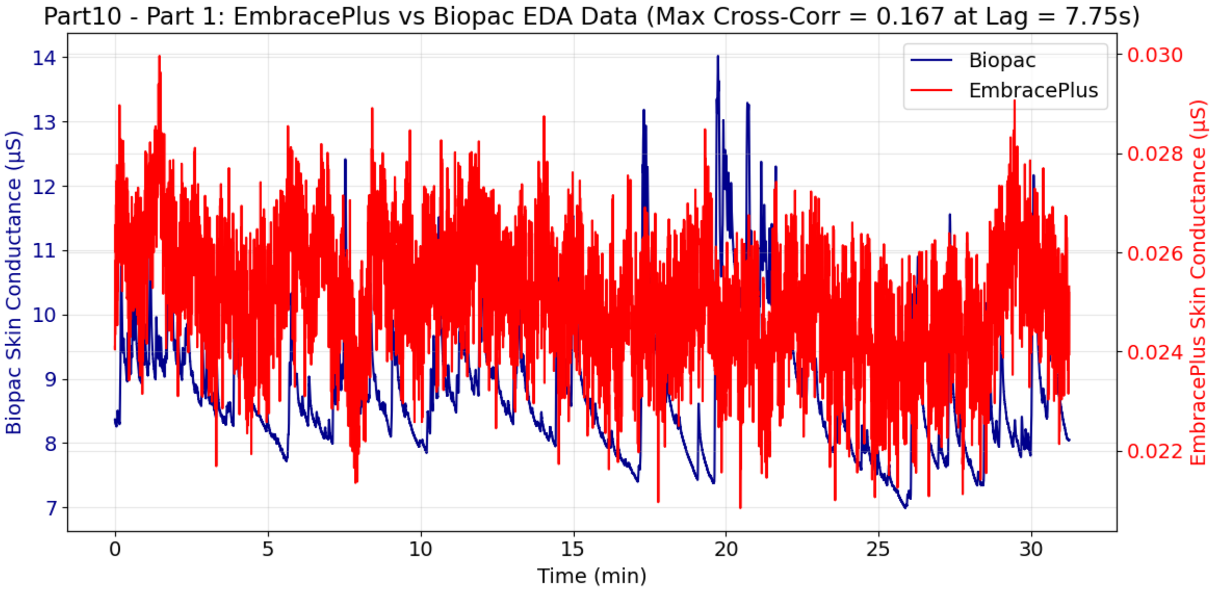

Part04: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.016 at Lag = 5.69s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
7
8
9
10
11
12
13
14
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
10
20
30
40
50
60
70

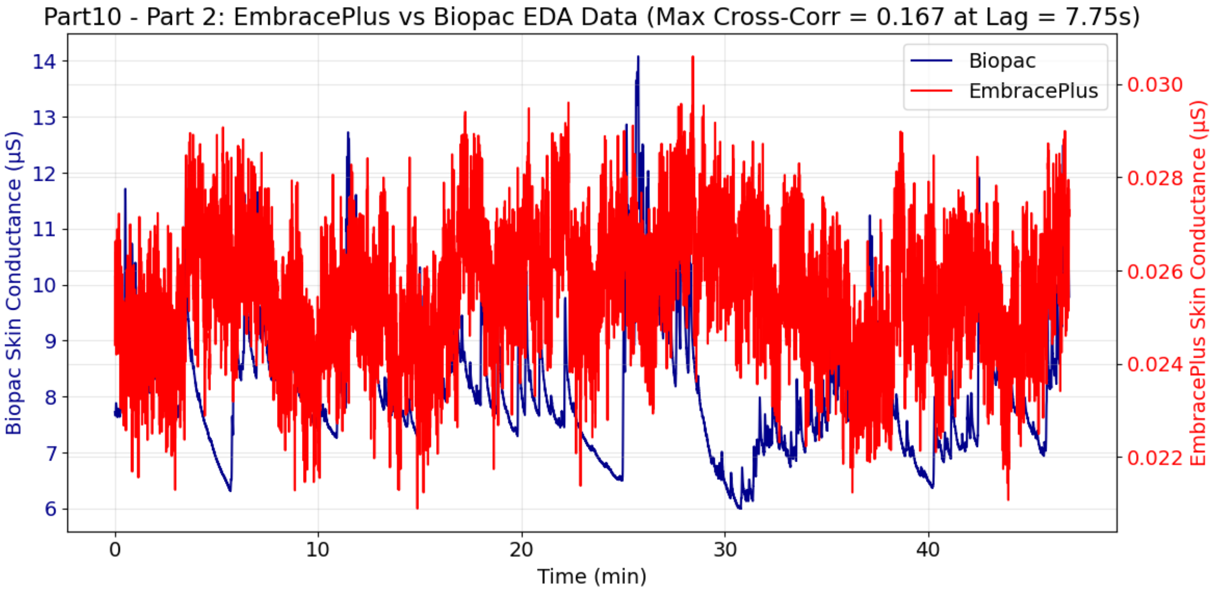

Part05: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.127 at Lag = -5.81s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
0
2
4
6
8
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
50
60
70

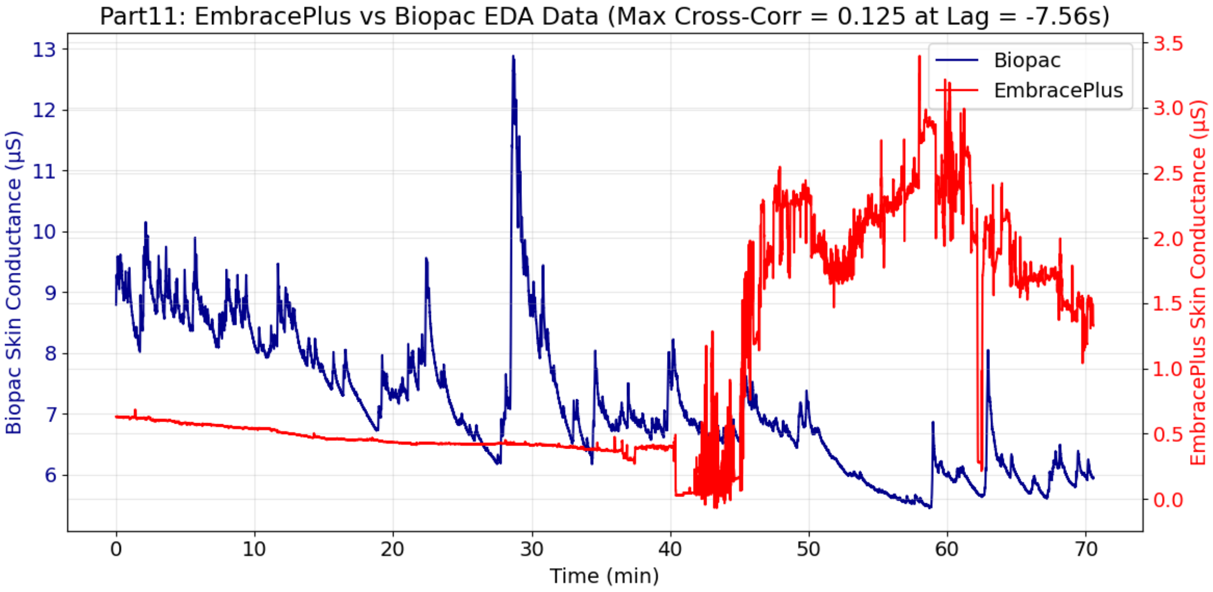

Part06: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.429 at Lag = -3.00s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
0
2
4
6
8
10
12
14
16
2
3
4
5
6
0
20
40
60
80

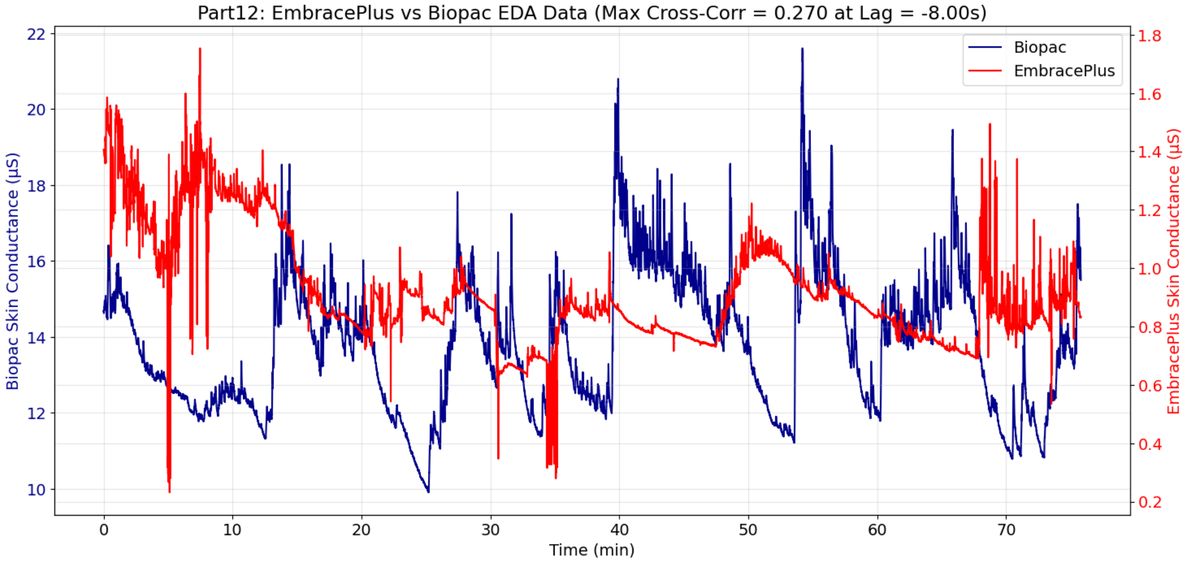

Part07: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.797 at Lag = 0.19s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
14
13
12
11
10
9
8
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0
10
20
30
40
50
60
70
80

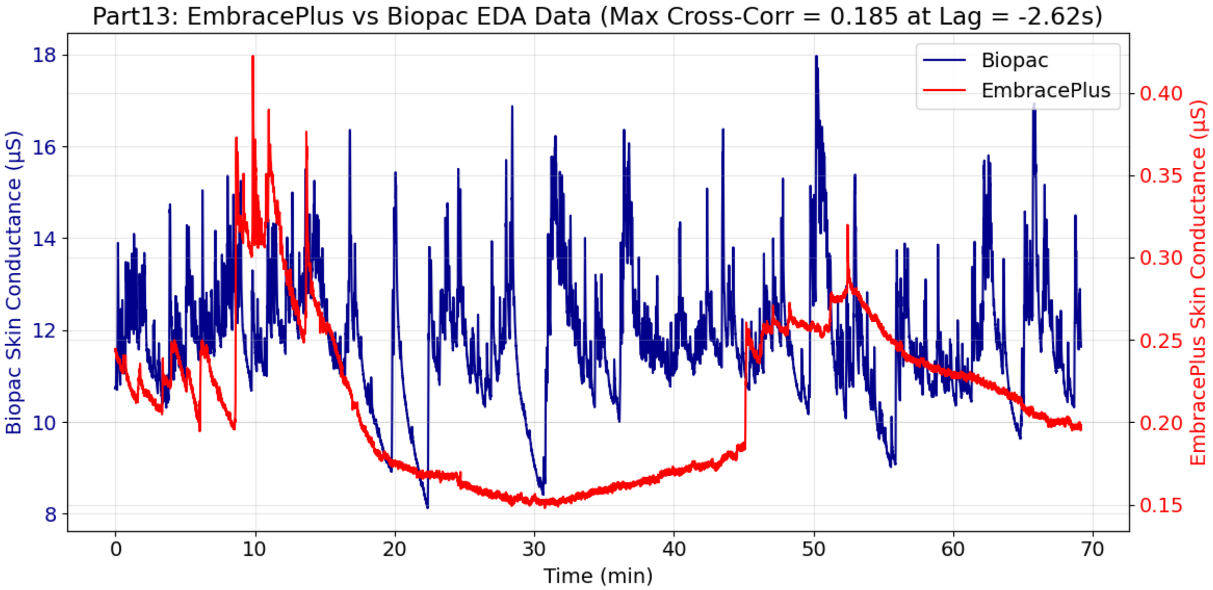

Part08: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.337 at Lag = -0.12s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
16
14
12
10
8
6
0.080
0.075
0.070
0.065
0.060
0.055
0.050
0.045
0
20
40
60
80

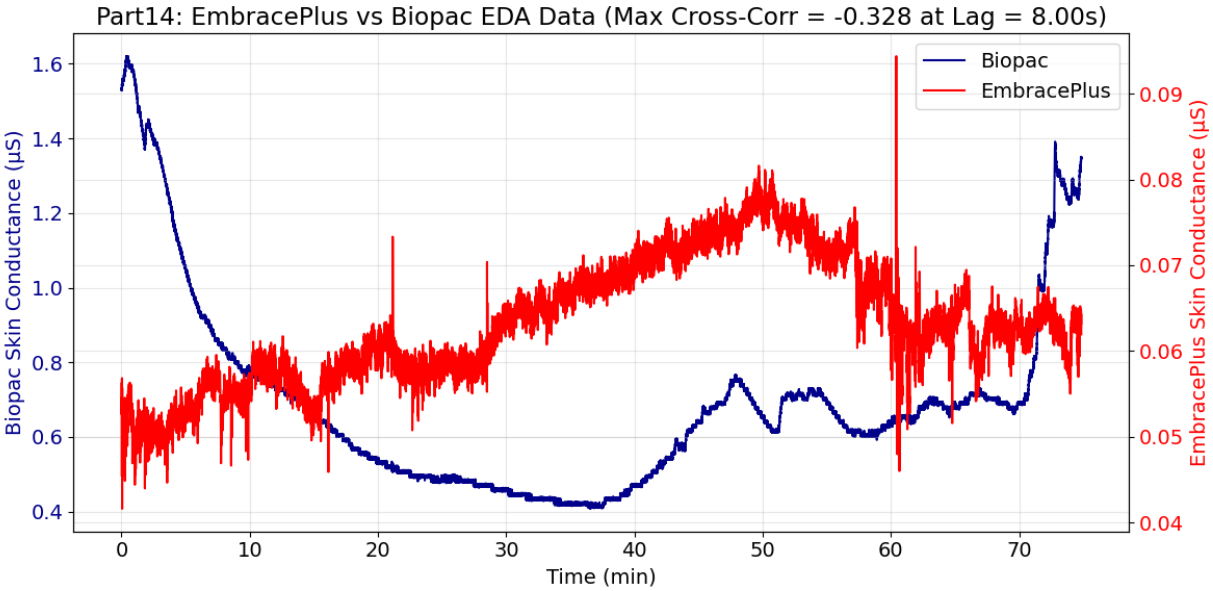

Part10 - Part 1: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.167 at Lag = 7.75s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
14
13
12
11
10
9
8
7
0.030
0.028
0.026
0.024
0.022
0
5
10
15
20
25
30

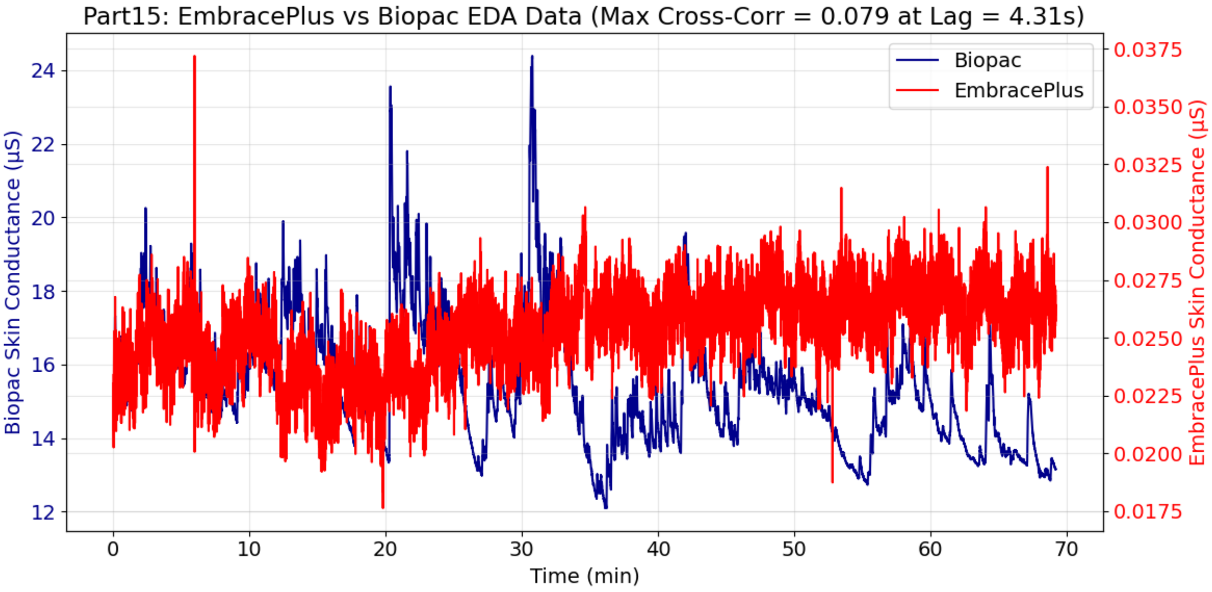

Part10 - Part 2: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.167 at Lag = 7.75s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
14
13
12
11
10
9
8
7
6
0.030
0.028
0.026
0.024
0.022
0
10
20
30
40

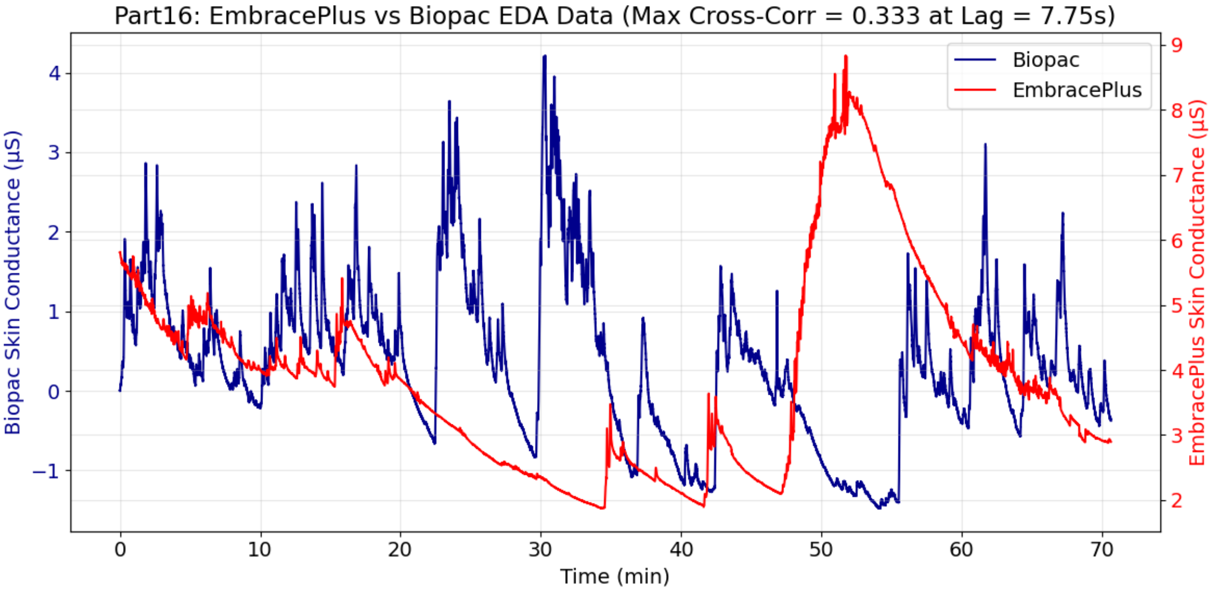

Part11: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.125 at Lag = -7.56s)

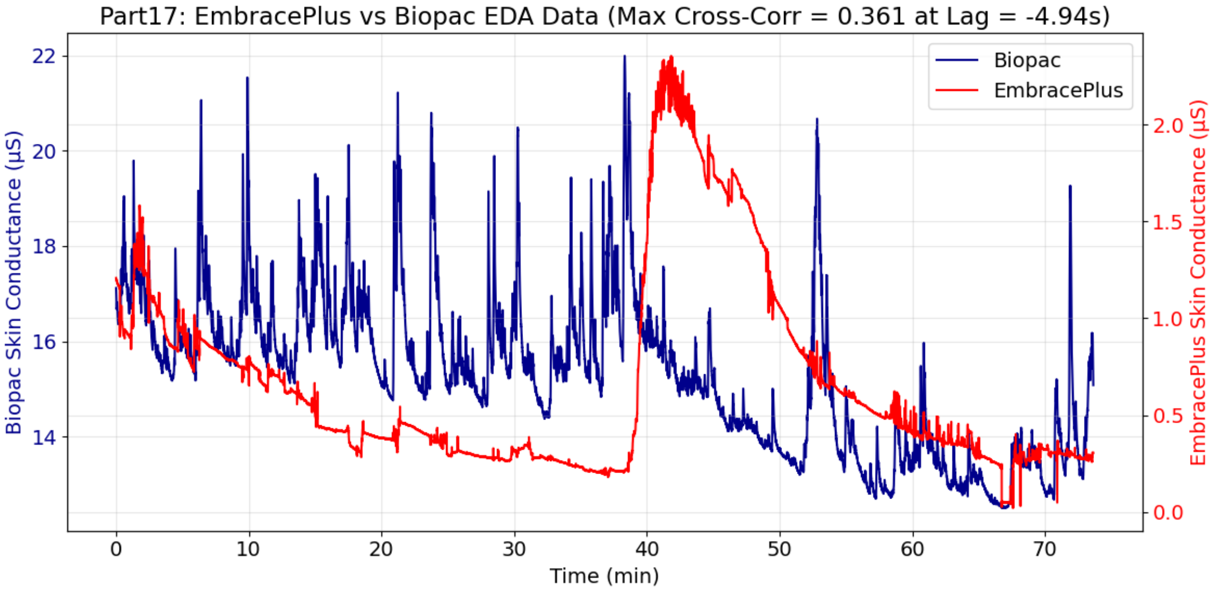

Part12: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.270 at Lag = -8.00s)

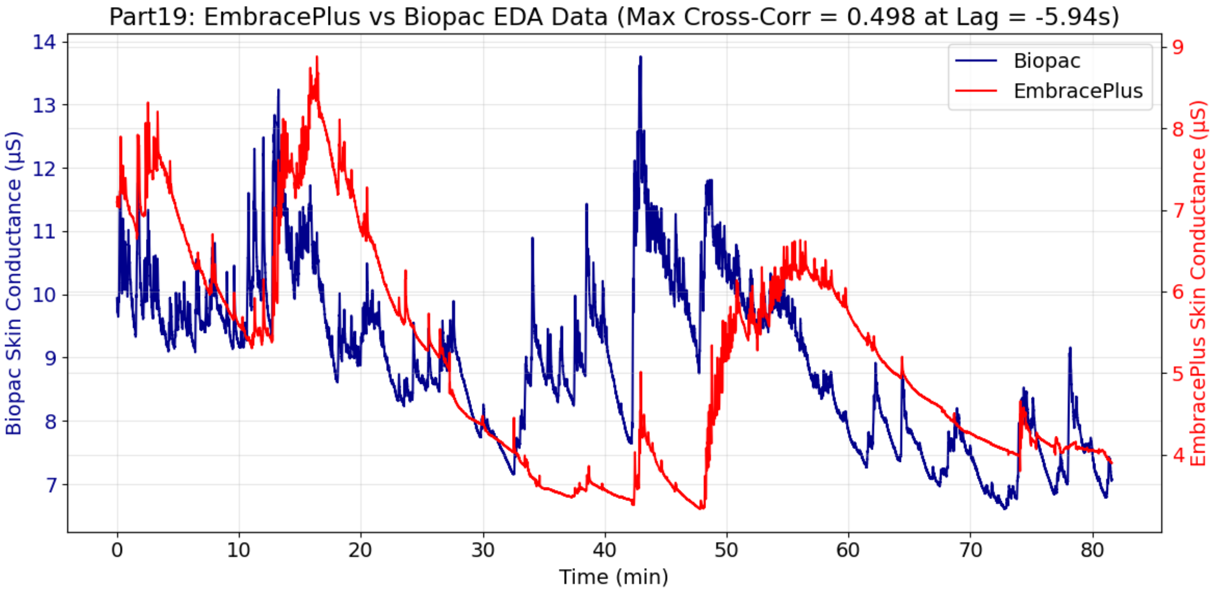

Part13: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.185 at Lag = -2.62s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
8
10
12
14
16
18
0.15
0.20
0.25
0.30
0.35
0.40
0
10
20
30
40
50
60
70

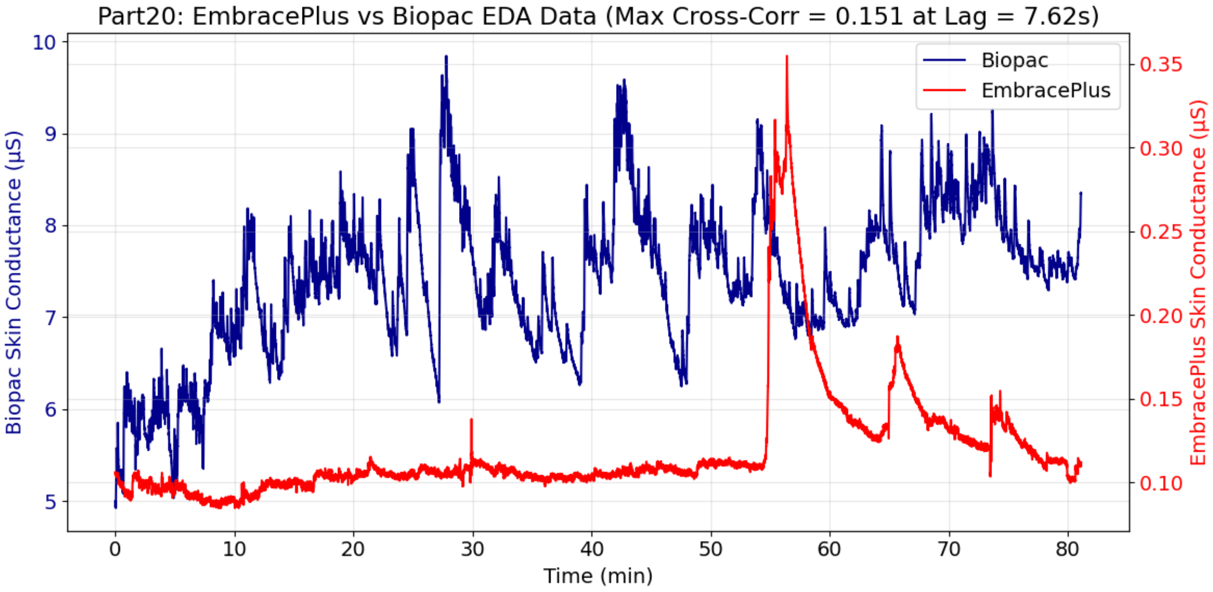

Part14: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = -0.328 at Lag = 8.00s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0.04
0.05
0.06
0.07
0.08
0.09
0
10
20
30
40
50
60
70


Part15: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.079 at Lag = 4.31s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
12
14
16
18
20
22
24
0.0175
0.0200
0.0225
0.0250
0.0275
0.0300
0.0325
0.0350
0.0375
0
10
20
30
40
50
60
70


Part16: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.333 at Lag = 7.75s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
−1
0
1
2
3
4
2
3
4
5
6
7
8
9
0
10
20
30
40
50
60
70


Part17: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.361 at Lag = -4.94s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
22
20
18
16
14
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
70


Part19: EmbracePlus vs Biopac EDA Data (Max Cross-Corr = 0.498 at Lag = -5.94s)
Biopac
EmbracePlus
Biopac Skin Conductance (µS)
EmbracePlus Skin Conductance (µS)
Time (min)
14
13
12
11
10
9
8
7
9
8
7
6
5
4
0
10
20
30
40
50
60
70
80



Presence Questionnaires

The questionnaire regarding Social Presence included following questions:

Q7_1: I felt like I was in the presence of another person in the virtual environment

Q7_2: I felt that the people in the virtual environment were aware of my presence

Q7_3: The people in the virtual environment appeared to be sentient (conscious and alive) to me

Q7_4: During the simulation there were times where the computer interface seemed to disappear, and I felt like I was working directly with another person

Q7_5: I had a sense that I was interacting with other people in the virtual environment, rather than a computer simulation

Response Distribution for Social Presence Questionnaire

Question	Completely Disagree	Disagree	Neither	Agree	Strongly Agree
Q7_1	0	4	4	8	0

Question	Completely Disagree	Disagree	Neither	Agree	Strongly Agree
Q7_2	0	1	3	11	1
Q7_3	3	6	5	2	0
Q7_4	3	8	1	4	0
Q7_5	4	2	4	6	0

Note: Response distribution for each item of the Social Presence Questionnaire (Q7_1 to Q7_5). Values represent the number of participants who selected each response option on a 5-point Likert scale ranging from Completely Disagree to Strongly Agree.

Physical Presence Questionnaire Results

The questionnaire regarding Physical Presence included following questions:

Q8_1: The virtual environment seemed real to me

Q8_2: I had a sense of acting in the virtual environment, rather than operating something from outside

Q8_3: My experience in the virtual environment seemed consistent with my experiences in the real world

Q8_4: While I was in the virtual environment, I had a sense of "being there"

Q8_5: I was completely captivated by the virtual world

Response Distribution for Physical Presence Questionnaire

Question	Completely Disagree	Disagree	Neither	Agree	Strongly Agree
Q8_1	1	6	1	8	0
Q8_2	0	3	3	9	1
Q8_3	1	5	6	4	0
Q8_4	0	2	3	11	0
Q8_5	1	7	5	2	1

Note: Response distribution for each item of the Physical Presence Questionnaire (Q8_1 to Q8_5). Values indicate the number of participants selecting each response on a 5-point Likert scale ranging from Completely Disagree to Strongly Agree.

Post-experiment questionnaire

Fear of Public Speaking

Public speaking in the VR TSST scenario was experienced consistently stressful, while not all participants reported fear, many experienced some level of discomfort or anxiety. For instance, *Participant 12* stated the feeling of being observed and judged. Moreover, *Participant 13* admitted of having the fear of running out of words while giving the speech, and *Participant 10* found it stressful due to time pressure, because there was not enough time to prepare properly for the speech. In addition, *Participant 15* mentioned feeling a bit of anxiety at the beginning of the job interview speech, and *Participant 17* remarked being slightly feared of public speaking and getting nervous when doing so. Furthermore, *Participant 14* described experiencing only *"medium arousal but not fear."* Notably, *Participant 07* differentiated based on context, meaning only being scared when there are a lot of people to present to. These reflections indicate that the social evaluation and task demands of the TSST were effective in eliciting some fear, even when participants were conscious of the artificiality of the environment.

Motion Sickness

Motion sickness was rarely reported, though minor discomfort surfaced occasionally. Most participants directly denied experiencing any symptoms: *"No motion sickness"* (Participant 12), *"No"* (Participant 05, 06, 07, 13). A few mentioned mild issues such as headaches or light dizziness. *Participant 08* noted, *"Yes; dizzy pre-speech,"* and *Participant 11* reported, *"Just a tiny bit 1/10." Participant 14* said, *"No motion sickness; slight headache,"* and *Participant 17* mentioned a *"2/10"* motion sickness experience during the butterfly task. These results highlight that while the VR system was generally well tolerated, minor physical discomforts still emerged, likely influenced by headset comfort, visual flow, or task pacing.

Nature Scenario and Relaxation

Responses to the nature scenario (Catching Butterflies) were diverse, ranging from relaxed enjoyment to frustration or even anxiety. Some participants appreciated the environment: *Participant 13* noted, *"Yes; nice environment,"* and *Participant 20* said, *"Yes; very relaxing; nice sounds."* Others, however, struggled with the goal-oriented structure, which distracted from any relaxing effect. *Participant 05* said that the environment is relaxing but that they were too focussed on the task of releasing the butterflies. Similarly, Participant *10* admitted not being relaxed due to the focus on the butterflies. Moreover, several participants experienced uncertainty or stress related to the task: *Participant 07* described a *"middle ground; uncertain task load,"* and *Participant 11* noted being *"stressed by task length."* Surprisingly, *Participant 18* reported that it was not relaxing and having anxiety due to the butterflies, and similar *Participant 15 was* feared of jump scares: *"No; afraid something jumps."*

Presence Experience

Most participants experienced moderate to strong physical presence, though social presence and realism were often disrupted. *Participant 10* said, *"Present; lights and sounds impactful,"* and *Participant 13* described *"strong elevator presence."* Others, however, noted immersion breaks: *Participant 07* observed that the *"judges table shift broke presence,"* and *Participant 11* coped by finding the situation humorous: *"Aware of VR; humorous coping."* Audio and visual glitches were commonly cited. *Participant 12* mentioned *"AI-sounding voices reduced realism,"* and *Participant 15* experienced *"light leakage in headset; robotic voices."* Interestingly, some specific design elements enhanced presence, for instance, *Participant 16* noted, *"Mosquito sounds enhanced presence."* These responses suggest that while physical immersion was largely achieved, the social and emotional believability of the scenarios often fell short due to technical constraints or scripting limitations.

Technical and Experimental Issues

Technical issues were widespread but manageable, affecting both the VR system and physiological sensors. Common VR problems included *Meta Quest 3 connectivity failures*, *controller malfunctions* (Participant 17), and *boundary misalignments* (Participant 04, 18). Sensor failures involved *ECG (Participant 05),* and *EDA disconnects (Participant 10)*, while *AcqKnowledge crashes* occurred for Participants 06 and 03. Some issues required alternate hardware (switching to Meta Quest 2). These findings emphasize the need for robust equipment protocols, regular calibration to ensure both data integrity and participant experience.

References

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The Trier Social Stress Test: Principles and practice. *Neurobiology of Stress*, *6*(1), 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001

Bareišytė, L., Slatman, S., Austin, J., Rosema, M., van Sintemaartensdijk, I., Watson, S., & Bode, C. (2024). Questionnaires for evaluating virtual reality: A systematic scoping review. *Computers in Human Behavior Reports*, *16*, 100505. https://doi.org/10.1016/j.chbr.2024.100505

BIOPAC Systems, Inc. (2024). *BIOPAC systems*. https://www.biopac.com/

Boucsein, W. (2012). *Electrodermal Activity*. Springer US. https://doi.org/10.1007/978-1-4614-1126-0

Braithwaite, J., Derrick, D., Watson, G., Jones, R., & Rowe, M. (2015). *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*. https://www.birmingham.ac.uk/Documents/college-les/psych/saal/guide-electrodermal-activity.pdf

Brunyé, T. T., & Giles, G. E. (2023). Methods for eliciting and measuring behavioral and physiological consequences of stress and uncertainty in virtual reality. *Frontiers in Virtual Reality*, *4*. https://doi.org/10.3389/frvir.2023.951435

Canali, S., Schiaffonati, V., & Aliverti, A. (2022). Challenges and Recommendations for Wearable Devices in Digital health: Data quality, interoperability, Health equity, Fairness. *PLOS Digital Health*, *1*(10). https://doi.org/10.1371/journal.pdig.0000104

Dhiman, P., Ma, J., Qi, C., Bullock, G. S., Sergeant, J. C., Riley, R. D., & Collins, G. S. (2023). Sample size requirements are not being considered in studies developing prediction models for binary outcomes: a systematic review. *BMC Medical Research Methodology*, *23*(1). https://doi.org/10.1186/s12874-023-02008-1

Fromberger, P., Meyer, S., Kempf, C., Jordan, K., & Müller, J. L. (2015). Virtual Viewing Time: The Relationship between Presence and Sexual Interest in Androphilic and Gynephilic Men. *PLOS ONE*, *10*(5), e0127156. https://doi.org/10.1371/journal.pone.0127156

Grassini, S., & Laumann, K. (2020). Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00349

Halbig, A., & Latoschik, M. E. (2021). A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality. *Frontiers in Virtual Reality*, *2*. https://doi.org/10.3389/frvir.2021.694567

Hongn, A., Bosch, F., Prado, L. E., José Manuel Ferrández, & Bonomini, M. P. (2025). Wearable Physiological Signals under Acute Stress and Exercise Conditions. *Scientific Data*, *12*(1). https://doi.org/10.1038/s41597-025-04845-9

Huang, Z., Xu, Y., Cheng, Y., Xue, M., Deng, M., Jaffrezic-Renault, N., & Guo, Z. (2022). Recent advances in skin-like wearable sensors: sensor design, health monitoring, and intelligent auxiliary. *Sensors & Diagnostics*, *1*(4), 686–708. https://doi.org/10.1039/D2SD00037G

Kim, K., Park, D.-Y., Song, Y. J., Han, S., & Kim, H. J. (2021). Consumer-grade sleep trackers are still not up to par compared to polysomnography. *Sleep and Breathing*. https://doi.org/10.1007/s11325-021-02493-y

Kooiman, T. J. M., Dontje, M. L., Sprenger, S. R., Krijnen, W. P., van der Schans, C. P., & de Groot, M. (2015). Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation*, *7*(1). https://doi.org/10.1186/s13102-015-0018-5

Kühne, C., Kecelioglu, E. D., Maltby, S., Hood, R. J., Knott, B., Ditton, E., Walker, F. R., & Kluge, M. G. (2023). Direct comparison of virtual reality and 2D delivery on sense of presence, emotional and physiological outcome measures. *Frontiers in Virtual Reality*, *4*. https://doi.org/10.3389/frvir.2023.1211001

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting. *Frontiers in Psychology*, *08*(213). https://doi.org/10.3389/fpsyg.2017.00213

Lemmens, J. S., Simon, M., & Sumter, S. R. (2021). Fear and loathing in VR: The emotional and physiological effects of immersive games. *Virtual Reality*, *26*. https://doi.org/10.1007/s10055-021-00555-w

Liu, J., Zhang, Y., Wang, Z., & Zhao, G. (2019). Validation of VR-based Immersive High-altitude Scene in Inducing Stress Response. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2257–2261. https://doi.org/10.1177/1071181319631036

Lu, Y., & Lau, S.-K. (2025). Ecological validity of VR experiments for psychological and physiological responses to audio-visual environments. *Building and Environment*, *274*, 112767. https://doi.org/10.1016/j.buildenv.2025.112767

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

McDuff, D., Thomson, S., Samy Abdel-Ghaffar, Galatzer-Levy, I. R., Poh, M.-Z., Sunshine, J., Barakat, A., Heneghan, C., & Sunden, L. (2024). What Does Large-scale Electrodermal Sensing Reveal? *BioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2024.02.22.581472

Meehan, M., Insko, B., Whitton, M., & Brooks, F. P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, *21*(3). https://doi.org/10.1145/566654.566630

Mehl, M. R., Eid, M., Wrzus, C., Harari, G. M., & Ebner-Priemer, U. W. (2023). *Mobile Sensing in Psychology*. Guilford Publications.

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

Milstein, N., & Gordon, I. (2020). Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States. *Frontiers in Behavioral Neuroscience*, *14*. https://doi.org/10.3389/fnbeh.2020.00148

Palombi, T., Galli, F., Giancamilli, F., D'Amico, M., Alivernini, F., Gallo, L., Neroni, P., Predazzi, M., De Pietro, G., Lucidi, F., Giordano, A., & Chirico, A. (2023). The role of sense of presence in expressing cognitive abilities in a virtual reality task: an initial validation study. *Scientific Reports*, *13*(1), 13396. https://doi.org/10.1038/s41598-023-40510-0

Pattyn, N., Pierre-François Migeotte, Neyt, X., van, & Cluydts, R. (2010). Comparing real-life and laboratory-induced stress reactivity on cardio-respiratory parameters: Differentiation of a tonic and a phasic component. *Physiology & Behavior*, *101*(2), 218–223. https://doi.org/10.1016/j.physbeh.2010.04.037

Peake, J. M., Kerr, G., & Sullivan, J. P. (2018). A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations. *Frontiers in Physiology*, *9*(9). https://doi.org/10.3389/fphys.2018.00743

Schoenmakers, M., Saygin, M., Sikora, M., Vaessen, T., Noordzij, M., & de Geus, E. (2025). Stress in action wearables database: A database of noninvasive wearable monitors with systematic technical, reliability, validity, and usability information. *Behavior Research Methods*, *57*(6). https://doi.org/10.3758/s13428-025-02685-4

Shakila Meshkat, Mahsa Edalatkhah, Corinna Di Luciano, Martin, J., Kaur, G., Gyu Hee Lee, Park, H., Torres, A., Mazalek, A., Kapralos, B., Dubrowski, A., & Bhat, V. (2024). Virtual Reality and Stress Management: A Systematic Review. *Cureus*. https://doi.org/10.7759/cureus.64573

Siavash Eftekharifar, Thaler, A., & Troje, N. F. (2021). Restorative Effects of Visual and Pictorial Spaces After Stress Induction in Virtual Reality. *ACM Symposium on Applied Perception*, 1–5. https://doi.org/10.1145/3474451.3476234

Siepmann, N., & Kowalczuk, J. K. (2021). Smart wearable devices for health and fitness: A scoping review. *Sensors*, *21*, 10.

Smyth, J. M., Sliwinski, M. J., Zawadzki, M. J., Scott, S. B., Conroy, D. E., Lanza, S. T., Marcusson-Clavertz, D., Kim, J., Stawski, R. S., Stoney, C. M., Buxton, O. M., Sciamanna, C. N., Green, P. M., & Almeida, D. M. (2018). Everyday stress response targets in the science of behavior change. *Behaviour Research and Therapy*, *101*, 20–29. https://doi.org/10.1016/j.brat.2017.09.009

Standard, W., Zielasko, D., Dawans, B. V., Domes, G., & Weyers, B. (2020). Open TSST VR: Trier Social Stress Test in Virtual Reality. *GI VR/AR Workshop 2020*. https://doi.org/10.18420/vrar2020_11

Stefano Triberti, Sapone, C., & Riva, G. (2025). Being there but where? Sense of presence theory for virtual reality applications. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-04380-3

Toczek, Y. (2018). *The influence of visual realism on the sense of presence in virtual environments*. https://pure.tue.nl/ws/files/46946153/855787-1.pdf

van Dooren, M., de Vries, J. J. G. (Gert-Jan), & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, *106*(2), 298–304. https://doi.org/10.1016/j.physbeh.2012.01.020

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. (2020). 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

Wang, D., Peng, Y., Haddouk, L., Vayatis, N., & Vidal, P.-P. (2025). Assessing virtual reality presence through physiological measures: a comprehensive review. *Frontiers in Virtual Reality*, *6*. https://doi.org/10.3389/frvir.2025.1530770

Wikipedia Contributors. (2019, October 27). *Wetland conservation*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Wetland_conservation

Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., & Sahebkar, A. (2017). The impact of stress on body function: A review. *EXCLI Journal*, *16*(1). PubMed. https://doi.org/10.17179/excli2017-480