Taylor & Francis
Taylor & Francis Group

Stress
The International Journal on the Biology of Stress

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ists20

Prior exposure to a sensorimotor game in virtual reality does not enhance stress reactivity toward the OpenTSST VR

Lena Sophie Pfeifer, Carina Zoellner, Oliver T. Wolf, Gregor Domes & Christian J. Merz

To cite this article: Lena Sophie Pfeifer, Carina Zoellner, Oliver T. Wolf, Gregor Domes & Christian J. Merz (2024) Prior exposure to a sensorimotor game in virtual reality does not enhance stress reactivity toward the OpenTSST VR, Stress, 27:1, 2361237, DOI: 10.1080/10253890.2024.2361237

To link to this article: https://doi.org/10.1080/10253890.2024.2361237

STRESS
2024, VOL. 27, NO. 1, 2361237
https://doi.org/10.1080/10253890.2024.2361237

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Prior exposure to a sensorimotor game in virtual reality does not enhance stress reactivity toward the OpenTSST VR

Lena Sophie Pfeifer[a*], Carina Zoellner[a*], Oliver T. Wolf[a], Gregor Domes[b,c] and Christian J. Merz[a]

[a]Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany; [b]Biological & Clinical Psychology, Trier University, Trier, Germany; [c]Institute for Cognitive and Affective Neuroscience, Trier University, Trier, Germany

ABSTRACT

Compared to the in-person Trier Social Stress Test (TSST), virtual reality (VR) variants reduce resource-intensity and improve standardization but induce stress with smaller effect sizes. However, higher cortisol reactivity is given for more immersive TSST-VRs. Immersivity depends on the VR-system, but perceived immersion may be targeted by exposure to, or interaction with the VR. We investigated whether stress reactivity towards the openly accessible OpenTSST VR can be enhanced by prior exposure to a sensorimotor game completed in VR as mediated by increased immersion. Therefore, $N=58$ healthy participants underwent the OpenTSST VR or its inbuilt control condition (placebo TSST-VR, pTSST-VR). Beforehand, participants completed a sensorimotor game either in VR or in real life. Stress was measured by means of self-reports, salivary cortisol concentrations, and salivary alpha-amylase (sAA) activity. Perceived immersion was assessed with the Igroup Presence Questionnaire (IPQ). The TSST-VR-group showed higher subjective stress than the pTSST-VR-group. Even though area under the curve measures indicated significant differences in cortisol levels between TSST-VR and pTSST-VR, this effect was not replicated in omnibus-analyses. Likewise, sAA was not responsive to stress. Our data suggests the OpenTSST VR does not reliably trigger physiological stress reactivity. Likewise, participants playing the VR-game before exposure to the TSST-VR did not show enhanced stress reactivity. Importantly, playing the VR-game did not lead to increased immersion (indicated by the IPQ), either. The key question resulting from our study is which manipulation may be fruitful to obtain a comparable stress response toward the TSST-VR compared to the in-person TSST.

ARTICLE HISTORY
Received 12 October 2023
Accepted 24 May 2024

KEYWORDS
Stress induction; immersion; immersivity; sensorimotor contingencies; presence

Introduction

The Trier Social Stress Test (TSST, Kirschbaum et al., 1993) can be considered a gold-standard in experimental stress induction methods (Allen et al., 2017). It combines key features of stressors with a special emphasis on social-evaluative threat (Dickerson & Kemeny, 2004) and has been shown to produce responder rates of 70-80% (Kudielka et al., 2007) in terms of cortisol reactivity which can be detected with a large effect size (d' = .925, Goodman et al., 2017). Moreover, the TSST leads to significant increases in salivary alpha-amylase (sAA) activity (Nater et al., 2005; 2006; Rohleder et al., 2004) which has been conceptualized as a marker of activation of the vegetative nervous system (Nater et al., 2005, 2006; Nater & Rohleder, 2009; Rohleder et al., 2004). With that, the TSST serves for standardized acute stress induction in the laboratory which might be applied in the context of diverse stress-related research questions (e.g., research questions concerning the nature or determinants of the stress response itself or research questions concerning stress-induced effects on cognition and emotion that may be assessed after TSST exposure).

During the TSST, participants perform a job interview and subtraction in front of a panel of researchers that keep a neutral appearance, and do not provide feedback. The TSST requires personnel in form of the panel. Moreover, the TSST varies across laboratories and participants (i.e., panel members vary) so that it is not perfectly standardized. For instance, laboratories dispose of a pool of research assistants who are trained to take the role of panel members during a TSST at different times during the week or at different times during the whole period of data collection. Therefore, different participants might be confronted with a varying TSST panel. Publications by Goodman et al. (2017) and Labuschagne et al. (2019) suggested that such factors can influence effect sizes of cortisol reactivity. Most prominently, the sex of participants and panel members was evaluated to be relevant in that female only panels have been shown to produce lowest cortisol effects while panels that include the opposite sex of the participant seem to be most effective.

To improve resource-intensity and standardization, the TSST has been translated to virtual reality (VR) as *TSST-VR* previously (e.g., Shiban et al., 2016). Currently, different versions

CONTACT Lena Sophie Pfeifer lena.pfeifer@rub.de Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany
*These authors share first-authorship

Supplemental data for this article can be accessed online at https://doi.org/10.1080/10253890.2024.2361237.

of TSST-VRs co-exist which differ in their degree of immersivity: TSST-VRs are conducted in different VRs (Cave Automatic Virtual Environments (CAVEs), head-mounted displays (HMDs), or screens), and either use virtual avatars or prerecorded videos of human researchers (for an overview over different TSST-VRs, see Helminen et al., 2019). A meta-analysis clarified that overall, the TSST-VR elicits a cortisol response (Helminen et al., 2019), while more immersive TSST-VRs lead to higher cortisol reactivity. Still, with medium effect sizes (Helminen et al., 2019), the TSST-VR consistently undercuts in-person applications of the TSST which provoke large cortisol effects (Goodman et al., 2017, but see Helminen et al., 2021). As a result, there seems to be a need for a well-immersive version of the TSST-VR that may be used by researchers around the world. Indeed, the application of various versions of TSST-VRs (of which some seem to be less immersive than others) might not only impair effect sizes in terms of cortisol reactivity, but also conflicts with the purpose of better standardization as compared to the in-person TSST. The OpenTSST VR (von Dawans et al., 2022) might counteract this shortcoming since it is an openly accessible version of the TSST-VR which might be used across different laboratories. During the OpenTSST VR, the original protocol of the TSST is highly conserved (preparation, speech and math in front of a non-responsive panel) but it is realized in VR where virtual avatars form the established social-evaluative panel. Indeed, a validation of the OpenTSST VR is currently available as a preprint (Linnig et al., 2024) which rendered mixed results on stress responsiveness (i.e., subjective stress, heart rate and cortisol) towards the procedure. Direct comparisons between TSST-VRs and in-person applications have only been addressed by few studies. Zimmer et al. (2019) exposed participants to the TSST, or the placebo-TSST (pTSST, Het et al., 2009), either in-person, or in VR, respectively. They reported comparable cortisol reactivity between the in-person TSST and the TSST-VR, a finding that was not reported by Shiban et al. (2016). While advantages of the TSST-VR lie in resource-efficiency and standardization, lowered cortisol reactivity and variability across laboratories represent targets for improvement.

According to Slater (2009), two illusions unfold in VR: The *place illusion* triggers feelings of being in the VR. The *plausibility illusion* provokes the impression that events in VR are actually happening. While immersion is evaluated by the user, it depends on immersivity of the technology providing *sensorimotor contingencies*. That is, *valid sensorimotor actions* are given as events occurring in VR lead to meaningful changes in sensory-perceptual experience of the user. *Valid effective actions* imply that operations executed by the user causally result in alterations in VR. As a result, Slater (2009) emphasized that above-mentioned illusions in VR can be considered a perceptual rather than a cognitive phenomenon. Nevertheless, users can learn how to perceive a system's *sensorimotor contingencies* to gate processing of virtual input. The place illusion might be achieved by means of attention or some sort of 'mental recreation' in that participants still know that they reside in VR but in that they just behave normally (Slater, 2009). Indeed, normal behavior implies movement and exploration of the VR - two aspects that contribute to the occurrence of the plausibility illusion by rendering subtle but meaningful correlations between internal, proprioceptive and external, VR-generated sensations (Slater, 2009; Slater et al., 1995). Several studies tried to manipulate the level of immersion in VR. For instance, Melo et al. (2016) let participants watch 360° videos as presented via an HMD for different time intervals. Interestingly, even though this procedure did not involve any movement, presence ratings as assessed by means of the Igroup Presence Questionnaire (IPQ, Schubert, 2003) increased with time for men while they decreased for women. This effect was not replicated by Lachlan and Krcmar (2011). The importance of movement in VR was finally emphasized by Usoh et al. (1999) and Slater et al. (1995) who showed that participants indicated place illusions to be stronger when they actually used their body to walk around or move in VR compared to conditions in which they manipulated a device (e.g., pressed a button) in order to move forward. This result was confirmed from another perspective in that the vividness of movement imagery was positively correlated with self-reported feelings of presence in VR as assessed by means of the IPQ (Ferrara et al., 2021).

Taken together, more immersive versions of the TSST-VR trigger higher cortisol reactivity (Helminen et al., 2019). Exposure to, and interaction with VR leads to increased perceived immersion through enhanced processing of virtual input (Slater, 2009). This may result in increased stress reactivity toward the TSST-VR. We hypothesized prior sensorimotor interaction with the VR leads to enhanced stress reactivity toward the OpenTSST VR.

Methods

Sample

Our final sample comprised $N=58$ (25 women) healthy participants, since an a priori sample size calculation in G*Power (Faul et al., 2007) recommended a minimal sample size of $N=40$ participants. In particular, for a repeated measures one-way analysis of variance (ANOVA) assuming within-between interactions, we entered the following parameters: effect size $f=0.325$ (Helminen et al., 2019), power = 0.95, $\alpha=0.05$, correlation between repeated measures = 0.3, non-sphericity correction = 1.

Inclusion and exclusion criteria were defined a priori. Participants were eligible being aged 18-35 years, normally weighted (body mass index (BMI): 18-27 kg/m^2), right-handed, and fluent in German language. Mental, neurological, or physical disorders, regular medication (all kinds of medications that were taken on a regular basis, especially such that are shown to affect sympathetic nervous system and hypothalamus-pituitary-adrenal axis (HPA) reactivity), acute psychosocial strain, smoking, or intake of other drugs were considered exclusion criteria while moderate regular alcohol consumption was accepted. Participants must not have completed a study on psychosocial stress before and psychology students must not be more advanced than the third semester. Women were required to take hormonal contraceptives. This criterion was chosen from a logistical perspective. First, the university course framing the current study was limited in time so that we aimed at facilitating

participant recruitment. That is, women of younger age (which were expected to dominate our target student population) take hormonal contraceptives frequently. Second, we aimed to avoid the logistical effort of aligning testing sessions with specific menstrual cycle phases as it is typically done with naturally cycling women. We still ensured that women were not tested during the pill off-phase. For the two to four weeks preceding the experiment, participants should not engage in activities that have been shown to disrupt basal cortisol reactivity. That is, participants must not donate blood (Hoogerwerf et al., 2015), travel with time shift (Doane et al., 2010), work in night shifts (Grosser et al., 2022), and report viral infections or colds (Rezai et al., 2022), vaccinations (Phillips et al., 2005), or extremely stressful situations (Kudielka et al., 2009). With respect to the VR, we aimed for naïve participants so that regular use of VR devices represented an exclusion criterion just as motion sickness or dizziness/nausea. Normal vision or corrected vision (but only with contact lenses) was acceptable for the HMD. Regular exposure to video games/gaming was assessed but did not constitute an exclusion criterion. We decided so since the inclusion of respective individuals would have provided the opportunity for explorative analyses to test whether regular exposure to video games may have functioned in the same sense as our game-manipulation was supposed to do. That is, gamers that should be habituated to game-related input might be primed to better process such kind of input and show better responsivity toward it – an effect that might have expanded to the VR. However, since our final sample did not contain a noteworthy proportion of gamers, we did not realize this analysis ultimately.

Data collection took place between 02/2023 and 04/2023 at research facilities of the Ruhr University Bochum, Germany. Data collection was realized in the context of a university course of the third and fourth B.Sc. psychology semester. During that course, students accompany a study from conceptualization to completion which also involves that students help with data collection in taking the role of experimenters. Indeed, the current study was specifically designed for this university course. The objective was to provide the students an interesting procedure with a relatively easy design. Indeed, virtual reality is a technology that younger generations show increased interest in. We hoped that this effect would also arise during recruitment so that individuals might be interested to participate in a study using VR. Moreover, the strength that the TSST-VR does not require a real panel was considered a further advantage facilitating the scheduling of testing sessions for enrolled students. As a result, students (in total: $N = 14$) primarily engaged in recruitment and testing of participants in the current study. In addition, since the same university course is realized at different departments of the Faculty of Psychology, it is considered that students themselves also participate in the studies of parallel courses. Therefore, recruitment of the current study was focused on fellow psychology students. Of note, it was not recorded whether experimenters and participants actually knew each other in the various pairings.

This study was approved by the local ethic committee of the Ruhr University Bochum, Germany and conducted in accordance with the Declaration of Helsinki. In the interest of open and reproducible science, experimental design and data analysis of the current study were preregistered in a project at the Open Science Framework (OSF) under the link https://osf.io/9cjqk. Raw data and final analysis scripts can also be retrieved from this OSF using the link https://osf.io/yhg8s.

Experimental design

The experimental design (see Figure 1) involved data collection taking place in the afternoon so that testing sessions started between 12:40 pm and 6:00 pm. Prior to the testing session, participants received written information about the procedure and background of the current study. That is, participants were aware that during the experiment, they would be exposed to a VR and that they would perform a short game as well as a further condition which might be moderately stressful or not stressful. Participants did not know to which of the conditions they were allocated so that all participants received the same information. Concerning the stressful condition, it was conveyed that the situation might be well faced during everyday life so that there was no reason for excitement or tension.

Having given informed consent, participants provided a first measure of stress and affect parameters (see sections "Saliva samples" and "Questionnaires" for further details). Subsequently, participants engaged in the sensorimotor game, either in real life (RL) or in VR, followed by a second assessment of stress and affect parameters. Notably, the duration of the game also served as an acclimatization period prior to stress induction. Then, participants were exposed to the OpenTSST VR in form of the TSST-VR or its inbuilt control condition, a placebo version of the TSST-VR (pTSST-VR). Afterwards, participants gave stress and affect measures for the third time and completed the Igroup Presence Questionnaire (IPQ) evaluating their experience in the TSST-VR or the pTSST-VR. Importantly, the IPQ was not given after the game-manipulation since only one half of participants completed the game in VR whereas the other half played the game in RL. As a result, only half of our sample size (i.e., participants having engaged in the VR-game) would have been applicable for completion of the IPQ after the game-manipulation. Moreover, we aimed at minimizing the delay between the game-manipulation and the subsequent exposure toward the (p)TSST-VR in order to preserve effects that may have been temporary. Indeed, a break between the game-manipulation and the subsequent exposure toward the (p)TSST-VR was already given by another assessment of stress and affect measures as mentioned above. In the following, participants underwent two waiting periods during which they engaged in mandala painting or sudoku. Waiting periods served for assessing later cortisol responses such as cortisol recovery with stress and affect measures four (after the first waiting period) and five (after the second waiting period), respectively. Finally, participants were debriefed and reimbursed.

Virtual environment

In the current study, the virtual environment was shown via an HMD by VIVE (i.e., VIVE Pro 2, HTC Corporation, Taoyuan City

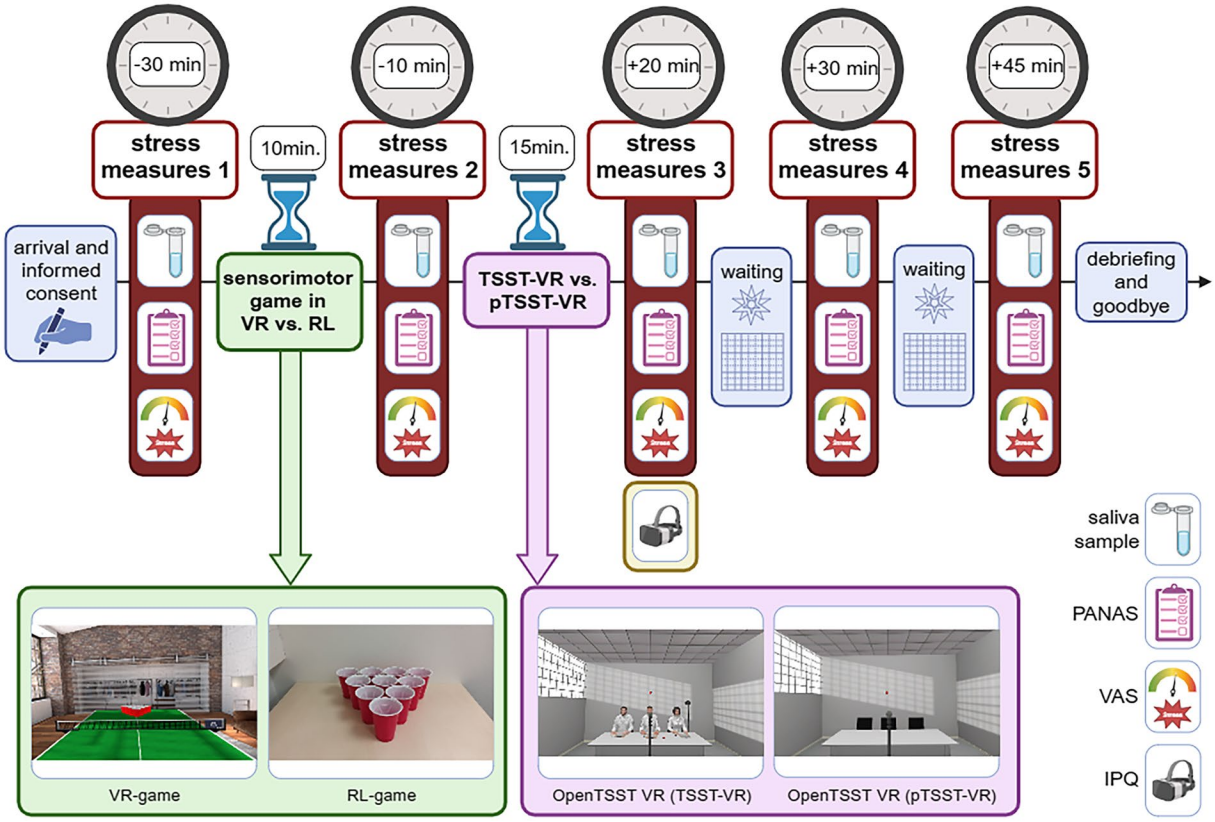

Figure 1. Overview over the experimental design of the current study.
Note. Stress measures covered saliva samples as well as self-report measures of stress and affect. Saliva samples were taken in form of Salivettes® (Sarstedt, Nümbrecht, Germany) and were analyzed for salivary cortisol concentrations and salivary alpha-amylase (sAA) activity. Subjective measures comprised the Positive and Negative Affect Schedule (PANAS) and a verbal analogue scale (VAS). For the VAS, participants were instructed to rate the stressfulness of the previous situation on a scale from 0 to 100. Timepoints of the stress and affect measures are given relative to stressor onset (i.e., onset of TSST-VR or pTSST-VR). On the bottom, there are screenshots/photos of the VR-game (VR = virtual reality), the RL-game (RL = real life), the TSST-VR, and the pTSST-VR, respectively. The VR-game was screenshotted from ©*ElevenTable Tennis*, By For Fun Labs, Inc., Austin, Texas. The Igroup Presence Questionnaire (IPQ) was given in parallel to our third assessment of stress and affect measures to capture perceived immersion of participants during the TSST-VR or pTSST-VR. Created with BioRender.com.

330, Taiwan) also disposing of inbuilt headphones. Moreover, for the VR version of the sensorimotor game, participants were equipped with controllers for their hands, also by VIVE. Importantly, experimenters assisted participants in putting on the HMD and in adjusting it to individual requirements (e.g., size of head, sharp vision). Likewise, experimenters helped to set down the HMD after completion of the TSST-VR or pTSST-VR. Of note, for participants who underwent the game in VR, there was no need to set down the HMD between the game and the TSST-VR or pTSST-VR. Instead, the second assessment of stress and affect measures was completed wearing the HMD in that experimenters placed the Salivette in the mouth of participants with gloves and read out the self-report measures while documenting the participants' responses. During the whole time spend in VR, participants were standing as this was required by the VR-game as well as by the TSST-VR or pTSST-VR. During the TSST-VR or pTSST-VR, we placed a real microphone in front of the participants.

Sensorimotor game

We manipulated experienced immersion using a two-group manipulation prior to stress induction. Participants were asked to play a sensorimotor game either in VR or in RL. For the VR-condition (VR-game), we used a mini-game included in *Eleven Table Tennis* (from ©*ElevenTable Tennis*, By For Fun Labs, Inc., Austin, Texas), which we acquired at Steam (Valve Corporation, Bellevue, WA). The chosen mini-game was called *Beer Pong* and was experienced from a first-person perspective. That is, participants stood in a virtual living room in front of a virtual table tennis table with ten virtual plastics cups on it. Virtual plastic cups were arranged in a pyramid form at the end of a virtual table tennis table. Furthermore, participants saw their hands as represented by the controllers which moved whenever participants moved in reality by engaging in the task. The game involved throwing a virtual ball into one of the ten virtual cups. The ball was continuously shown in VR so that participants could also follow its trajectory as thrown to the cups. Per round, players had 1 min to throw as many balls as they needed to strike all cups. Importantly, participants did not have to pick up the ball, but it always reappeared in their hands after each throw. For the RL-condition (RL-game), the set-up mimicked the one from the VR-game, though participants threw real table tennis balls into real plastic cups, which were half-filled with water to remain stable on a regular table. When all balls were thrown, participants recollected them and threw again. Both groups played the game for the duration of 10 min. For a screenshot of the VR-game and a photo of the RL-setup, see Figure 1.

We opted for this VR-game since it seemed to capture the features exposed to be relevant for creating the place and

the plausibility illusion in VR according to Slater (2009). That is, the VR-game required movement (i.e., grasping and throwing) of participants. On the one hand, movement was shown to contribute to increased reports of place illusion (Slater, 2009). On the other hand, we deemed the VR-game to cover *valid effective actions* in that there was a direct cause-effect relationship between throwing the virtual ball, hitting a virtual plastic cup and seeing the virtual cup disappear after being hit. As a result, we considered the chosen VR-game appropriate for increasing our participants' immersion into the VR. Moreover, we were in search of a VR-game that was equally well applicable in a real-world setting. Indeed, this aspect was also given with the current VR-game.

Stress induction: OpenTSST VR

For stress induction in VR, we applied the OpenTSST VR (von Dawans et al., 2022, http://www.uni-trier.de/index.php?id=88863), an openly accessible tool that is currently being validated (see preprint by Linnig et al., 2024). The OpenTSST VR can be downloaded autonomously or provided by Trier University via personal communication. The OpenTSST VR provides a virtual environment as generated with Steam VR as well as a graphical user interface (GUI) which allows to control over the protocol of the OpenTSST VR as well as the activation of virtual avatars during execution. The OpenTSST VR may be easily installed by unzipping the folders and copying them to a desired directory on a client PC. This client PC needs to be connected with an HMD device which is then inputted by the PC to setup the VR. In the current study, the OpenTSST VR was realized in a laboratory room which was slightly larger sized in order to provide enough space to allow for movement in the VR.

In the OpenTSST VR, the original TSST procedure is highly conserved. That is, after a preparation period of 5 min, participants apply for their dream job in front of a VR-animated panel for 5 min. Thereafter, participants have to serially subtract the number 13, starting at 2023, also for 5 min. Overall, the OpenTSST VR takes about 15 min. Importantly, instructions during the OpenTSST VR are given in written form (i.e., are depicted in the visual field in VR) and are also read out aloud to the participant via headphones.

Similar to the TSST as originally introduced by Kirschbaum et al. (1993), the VR-animated panel consists of three entities. That is, one female and two male researchers in white lab coats. During the preparation phase, the panel is still absent but suddenly appears with the start of the job interview. The panel is positioned at a table in a neutral room that resembles a common laboratory testing environment and that is animated in a 360° angle. Behind the panel, there is a video-camera creating the impression that the scenario is video-taped. Likewise, a microphone is placed right in front of the participant pretending that the tasks are audio-recorded. Importantly, in the current study, we further placed a real microphone in front of the participants. During the tasks, panel members are animated to occasionally make use of pen and paper in front of them, pretending they are taking notes. For the speech part, it is possible to further activate the avatars to voice statements such as "You still have time. Please continue." or "What are your strengths?". For the serial subtraction, avatars can be activated to express that participants made a mistake and must restart at 2023 (i.e., "That was wrong. Please start over at 2023.").

In the current study, the start of the OpenTSST VR as well as the activation of the panel members during the TSST was executed by a neutral experimenter. This was done via the GUI of the OpenTSST VR which allows to choose from various statements the avatars may voice. All experimenters were carefully trained to activate the virtual avatars whenever participants began to hesitate during the speech part and whenever participants made a mistake during the math part. Indeed, a comprehensive list of statements the virtual avatars are able to voice through the experimenter's activation can be found in the user's guide of the OpenTSST VR (https://www.uni-trier.de/fileadmin/fb1/prof/PSB/TKP/TSST-VR/Open_TSST_VR_User_s_Guide_20221123.pdf). We specified no fixed number of activations that must have been voiced. Indeed, we considered the activation of avatars as a tool that serves to provide individuality (thereby potentially diminishing standardization) and interactivity - two criteria that we deemed to be captured only without strict activation targets.

During exposure to the OpenTSST VR, the experimenter was sitting behind a curtain (approximately 1 m away from the participants) and kept silent while participants were instructed to follow requests made by the virtual panel. Importantly, the experimenter executed actions in the VR via a silent mouse to prevent the participants' awareness that the experimenter was actually controlling the avatars. For further details on the OpenTSST VR, please check out the official users' guide on the homepage (http://www.uni-trier.de/index.php?id=88863).

As a control condition, participants were exposed to the inbuilt placebo-version offered by the OpenTSST VR which is a VR-adaptation of the established pTSST (Het et al., 2009). Here, after a 5 min preparation period, participants are allowed to speak about their preferred topic for 5 min. Then, participants are instructed to count upwards from "0" in steps of "15". Both tasks are performed in the same virtual environment (i.e., the testing room) in front of an empty table while the panel is absent. Importantly, the microphone was still present during the pTSST-VR while the camera was not there.

Of note, we struggled with technical difficulties during testing sessions of $N = 18$ participants. In detail, these technical difficulties involved the curiosity that the VR-environment was no longer properly aligned within the real world. For instance, when starting the TSST-VR or pTSST-VR, participants found themselves standing in the floor whereas the table with the virtual panel was at the height of their chest. Indeed, these technical difficulties were solved by means of a recalibration of the virtual environment. This sometimes involved a delay of several minutes as well as participants to set down the HMD. Moreover, even though recalibration was required only once per every of those $N = 18$ participants, exact timepoints differed unsystematically across participants. Importantly, we repeated all analyses in a subsample of participants without technical difficulties (see Tabel S10-S13 and Figure S2 in the "Supplementary Material" for results on these analyses). Indeed, in this subsample, findings did not deviate from the pattern of results that was observed in the whole sample.

Saliva samples

We collected salivary cortisol levels and sAA activity by means of Salivettes® (Sarstedt, Nümbrecht, Germany) at different timepoints of the experiment (T1: −30 min, T2: −10 min, T3: +20 min, T4: +30 min, T5: +45 min, all times relative to stressor onset, Figure 1). We did not preregister the collection and analysis of sAA but evaluated it in order to obtain a more thorough view of physiological stress responses, including a measure of sympathetic activity. Salivettes were stored at −20 °C until data collection was completed. Subsequent analysis took place at the joint laboratory of the Genetic Psychology and the Cognitive Psychology departments of the Ruhr University Bochum. Cortisol was analyzed using commercially available enzyme-linked immunosorbent assays (ELISA; Demeditec, Kiel, Germany) while sAA activity was quantified as described elsewhere (Lorentz et al., 1999). All samples were analyzed in duplicates.

Analysis in duplicates was not possible for $N = 2$ samples so that we relied on single determination in these cases. $N = 3$ samples did not provide enough content for a valid analysis so that these values are missing. In total, we were able to include data of $N = 287$ saliva samples. Intra-assay coefficients of variations (CVs) were below 6.5% (cortisol) and 6% (sAA) and inter-assay CVs below 7% (cortisol) and 8% (sAA).

Questionnaires

In parallel to the acquisition of objective stress measures (i.e., saliva samples), we asked participants to respond to all items of the Positive and Negative Affect Scale (PANAS, German version: Krohne et al., 1996), which aims to assess subjective-emotional affect (divided for positive affect and negative affect) in a specific moment. In its German version, the PANAS was shown to reach Cronbach's alpha of .86 for both the positive and the negative affect scale (Breyer & Bluemke, 2016). Of note, we only analyzed the negative affect scale of the PANAS for the current publication since we considered the positive affect less important in the context of an acute stress induction. Additionally, we asked participants to respond to the question "How stressed were you feeling since the last time you responded to this question?"/"- the start of the experiment?" on a verbal analogue scale (VAS) ranging from 0 to 100. Both subjective measures were acquired verbally during saliva collection. At the start of the experiment, participants were presented with the instructions for the PANAS and the VAS and were asked to familiarize themselves with all items. We applied responses verbally since across the experiment, participants were partly wearing the HMD while assessing stress parameters.

After exposure to the TSST-VR or pTSST-VR and the third assessment of stress and affect measures (at +20 min), participants additionally completed the Igroup Presence Questionnaire (IPQ, German version: Schubert, 2003). Using 14 items, the IPQ measures the subjective sense of *spatial presence*, *involvement*, *experienced realism* and *sense of being there* after exposure to a virtual environment. In two independent German samples (study 1: $N = 264$, study 2: $N = 296$), the IPQ was confirmed to feature good validity and internal consistency (Regenbrecht & Schubert, 2002). For instance, Cronbach's alpha was reported to range between α = .85 (study 1) and α = .87 (study 2) for the whole 14-item questionnaire. The different subscales showed Cronbach's alpha as following: *spatial presence:* α = .80 (study 1) and α = .77 (study 2), involvement: α = .76 (study 1 and 2), *experienced realism*: α = .68 (study 1) and α = .70 (study 2). We expected the IPQ to reflect the game-manipulation of our study especially for the presence subscale. Indeed, several studies (e.g., Ferrara et al., 2021; Melo et al., 2016) found effects in the presence scale of the IPQ after movement interventions in VR. Indeed, we hypothesized this increased presence or increased immersion to transfer to the following (p)TSST-VR.

Data preparation

We prepared and summarized acquired data using Python (version 3.11.4) implemented in Spyder (version 5.4.2). All analyses were conducted using R (version 4.2.2) implemented in RStudio (RStudio Team, 2021). For analyses and plotting, we used the following R-packages: psych (Revelle, 2023), dplyr (Wickham et al., 2022), tidyr (Wickham & Girlich, 2022), rstatix (Kassambara, 2022b), ggplot2 (Wickham, 2016), ggpubr (Kassambara, 2022a), effectsize (Ben-Shachar et al., 2020), ARTool (Elkin et al., 2021), multcomp (Hothorn et al., 2008), rcompanion (Mangiafico, 2023), and DescTools (Andri et al., 2022).

Statistical analysis

Prior to statistical analysis, we checked the normality distribution of data by means of visual inspection of QQ-plots and by means of Shapiro-Wilk tests (of note, we deviated from our preregistration in omitting an outlier exclusion since this was suggested by a reviewer). When normality was violated, for (unaggregated) physiological stress measures (i.e., cortisol or sAA), we applied transformation by means of the natural logarithm and used parametric one-way analysis of variance (ANOVA) subsequently. When normality was violated for subjective measures (i.e., PANAS, VAS, IPQ) as well as for aggregated stress and affect measures (see below), we used non-parametric aligned rank transform (ART)-ANOVA. In both cases, we used Type II SSs and accounted for main effects as well as for two-way and three-way interactions. We checked for differences between our experimental groups in terms of demographic sample characteristics using ANOVA or Pearson's Chi-squared tests for categorial variables. For all analyses, we used the standard significance level of 0.05. and corrected for multiple comparisons with Holm-correction. Pairwise post-hoc tests were performed for significant main and interaction effects either using parametric t-tests or non-parametric post-hoc tests of the ART-package, again using Holm-correction. Effect sizes are given as partial eta squared (η_p^2).

For our hypotheses-driven analyses, we set up 2 (game-condition) x 2 (TSST-condition) x 5 (time) ANOVA which were run for all our dependent stress and affect measures separately (i.e., negative affect as given by the PANAS, self-reported stressfulness as given by the VAS, cortisol levels, and sAA activity). In contrast, the factor time was omitted for

the IPQ since this questionnaire was only given once as well as for analyses of repeated stress and affect measures in terms of aggregated indices. That is, all repeated stress and affect measures were additionally analyzed using the (1) area-under-the-curve with respect to increase (AUCi; Pruessner et al., 2003) as well as using (2) min-max differences (Miller et al., 2018). Importantly, for all dependent variables, these indices, (AUCi and min-max differences) were calculated using raw, untransformed data.

1. The AUCi reflects changes across time and indicates sensitivity of the system being studied. For AUCi measures, we included timepoints T2-T5 (T2: -10 min, T3: +20 min, T4: +30 min, T5: +45 min). This was done since timepoint T1 (at -30 min) was originally added to enable evaluations of whether the game-manipulation was stressful (i.e., by comparing T1 and T2). However, T2 was considered the proper baseline for analyses concerning stress reactivity, which is not confounded by the effect of the game-manipulation and anticipatory stress. We calculated AUCi measures based on the AUC with respect to ground (AUCg). That is, for each dependent variable, the AUCi was calculated in subtracting the product of the raw value of the dependent variable at the first timepoint (T2, -10 min.) and the added temporal distance between all timepoints (T2-T5) from the AUCg measures of the respective dependent variable.
2. Min-max differences acknowledge the fact that cortisol reactivity might be better captured by means of cortisol concentrations as observed after stressor onset compared to indexing a single baseline as it is done during AUCi analyses. Therefore, in addition to the AUCi, we identified the raw individual minimal cortisol concentration and the raw maximal cortisol concentration across the timepoints T2-T5 for each participant. Indeed, in the publication by Miller et al. (2018), min-max differences were created using individual minimal and maximal cortisol concentrations from timepoint "0" on. Since in our study, there was no assessment directly at stressor onset, we included T2 to approximate some sort of timepoint" 0", also in line with our AUCi approach. To form min-max differences, we then subtracted the raw individual minimal concentration from the raw maximal cortisol concentrations to create min-max differences.

Deviating from our preregistration, we also applied sex-dependent analyses for cortisol data. We did so since women have been shown to respond less strongly than men in terms of cortisol in stress induction protocols (see Liu et al., 2017 for the in-person TSST, and Helminen et al., 2019 and Santl et al., 2019 for the TSST-VR). Noteworthy, hormonal contraceptives can explain a substantial amount of lowered cortisol reactivity towards experimental stressors in women (Liu et al., 2017). This is of relevance since we tested women taking oral contraceptives in the current study. Concerning our analyses, we realized sex-specific analyses by adding sex as a factor to our repeated measures ANOVA so that it included both group-manipulations (game-condition and TSST-condition), sex and time as predictors for cortisol as our dependent variable. Results on this analysis can be found in the "Supplementary Material" (Table S9 and Figure S1).

Results

Sample

Our final sample comprised $N = 58$ participants. Demographic characteristics of the final sample can be found in Table 1. The groups were statistically equal concerning sex, age, BMI, overall gaming activity, regular gaming hours, and highest educational degree (all $p > 0.05$).

Immersion

No subscale of the IPQ revealed a significant effect of group affiliation, namely neither a main effect of game-condition, nor a main effect of TSST-condition (all $p_{Holm} > 0.05$). In line with that, we did not find an interaction effect between game- and TSST-condition for any of the IPQ subscales (all $p_{Holm} > 0.05$).

Subjective stress

Subjective stress is illustrated in Figure 2, panel A and B. Exhaustive statistical parameters can be found in detail in the "Supplementary Material".

Negative affect

We found a significant main effect of TSST-condition ($F_{(1, 53.954)} = 19.198$, $p_{Holm} < 0.001$, $\eta_p^2 = 0.262$) in that negative affect was significantly higher during the TSST-VR compared to the pTSST-VR across timepoints. The main effect of game-condition did not reach statistical significance ($p_{Holm} > 0.999$, see Table S1). However, there was a significant main effect of time ($F_{(4, 208.718)} = 45.148$, $p_{Holm} < 0.001$, $\eta_p^2 = 0.464$) as well as a significant interaction effect of time and TSST-condition ($F_{(4, 208.853)} = 12.572$, $p_{Holm} < 0.001$, $\eta_p^2 = 0.194$) whereas other interactions did not reach statistical significance (all $p_{Holm} > 0.05$, see Table S1). For the main effect of time, pairwise post-hoc comparisons revealed significant differences in negative affect across the TSST-conditions between various points in time (see Table S2). In line with that, pairwise post-hoc comparisons that concerned the interaction of time and TSST-condition indicated significant differences between several timepoints within the TSST-VR and the pTSST-VR (see Table S3). Likewise, several timepoints differed between TSST-VR and pTSST-VR (see Table S3). However, concerning the same point in time, significant differences between the TSST-VR and the pTSST-VR were only found at +20 min ($p_{Holm} < 0.01$) but at no other point in time (all $p_{Holm} > 0.05$). In detail, negative affect was significantly higher in the TSST-VR-group compared to the pTSST-VR at +20 min. Additional analyses of aggregated measures (AUCi and min-max differences) extended these results in revealing a main effect of TSST-condition (AUCi: $F_{(1, 54)} =$

Table 1. Demographic sample characteristics and descriptive aggregated stress and affect measures.

	TSST-VR		pTSST-VR		total
	VR-game	RL-game	VR-game	RL-game	
N (male:female)	15 (9:6)	15 (7:8)	14 (10:4)	14 (7:7)	58 (33:25)
Age (years)	22.400 (*2.667*)	21.200 (*2.651*)	22.286 (*3.338*)	21.714 (*1.939*)	21.897 (*2.667*)
BMI (kg/m^2)	22.476 (*1.819*)	22.371 (*2.483*)	23.284 (*2.088*)	23.400 (*3.160*)	22.867 (*2.412*)
Video-game consumption (yes:no)	5:10	5:10	8:6	5:9	22:35
AUCi Negative affect (PANAS)	15.233 (*16.356*)	11.183 (*11.993*)	2.732 (*8.733*)	1.875 (*9.560*)	7.944 (*13.105*)
Min-max Negative affect (PANAS)	1.027 (*0.596*)	0.947 (*0.782*)	0.536 (*0.438*)	0.436 (*0.553*)	0.745 (*0.645*)
AUCi Self-reported stressfulness (VAS)	705.833 (*674.45*)	686.000 (*510.94*)	12.857 (*487.94*)	337.857 (*552.18*)	444.61 (*617.71*)
Min-max Self-reported stressfulness (VAS)	55.000 (*22.162*)	52.800 (*25.109*)	33.286 (*23.627*)	29.643 (*20.875*)	43.069 (*25.146*)
AUCi Cortisol	35.712 (*106.986*)	11.268 (*55.365*)	−34.398 (*48.373*)	−16.592 (*59.955*)	−0.068 (*75.642*)
Min-max Cortisol	2.436 (*2.855*)	1.550 (*1.609*)	2.249 (*1.471*)	2.006 (*1.366*)	2.058 (*1.916*)
AUCi sAA	−158.82 (*3625.0*)	2999.01 (*7266.9*)	−266.63 (*4353.1*)	−888.48 (*4792.5*)	492.61 (*5324.3*)
Min-max sAA	124.79 (*102.49*)	195.92 (*319.19*)	111.08 (*113.55*)	143.50 (*136.41*)	144.39 (*189.37*)

Note. *N* (male:female) and mean (*standard deviation*) for demographic sample characteristics as well as for aggregated stress and affect measures (AUCi measures and min-max differences) separated for the four experimental groups (1) TSST-VR + VR-game, (2) TSST-VR + RL-game, (3) pTSST-VR + VR-game, and (4) pTSST-VR + RL-game, and for the whole sample (total). AUCi measures were calculated according to the formula by Pruessner et al. (2003) whereas min-max differences rely on a publication by Miller et al. (2018). Importantly, AUCi measures and min-max differences were calculated using raw, non-transformed data.

11.71, $p_{Holm} < 0.01$, $\eta_p^2 = 0.18$, min-max differences: $F_{(1, 48)} = 12.976$, $p_{Holm} < 0.01$, $\eta_p^2 = 0.213$) as negative affect was significantly higher during the TSST-VR compared to the pTSST-VR (AUCi: $p_{Holm} < 0.01$, min-max differences: $p_{Holm} < 0.001$).

Visual analogue scale

Analysis of VAS-ratings revealed a significant main effect of time ($F_{(4, 208.769)} = 69.273$, $p_{Holm} < 0.001$, $\eta_p^2 = 0.570$) as well as significant interaction effect of time and TSST-condition ($F_{(4, 208.844)} = 10.338$, $p_{Holm} < 0.001$, $\eta_p^2 = 0.165$). However (see Table S4), there was no significant main effect of TSST-condition (p_{Holm}= 0.087) and no significant main effect of game-condition ($p_{Holm} > 0.999$). Concerning the main effect of time, pairwise post-hoc comparisons revealed significant differences in VAS ratings between several points in time across the TSST-VR and the pTSST-VR (see Table S5). This was again confirmed by pairwise post-hoc comparisons unraveling the interaction in that time-dependent effects were found within the different groups of the TSST-condition (see Table S6). Likewise, TSST-VR and pTSST-VR differed from each other at various points in time (see Table S6), but a significant difference between the TSST-conditions was not found at one and the same point in time (all $p_{Holm} > 0.05$). Analyses of aggregated measures (AUCi and mix-max differences) further confirmed the main effect of TSST-condition (AUCi: $F_{(1, 54)} = 12.41$, $p_{Holm} < 0.01$, $\eta_p^2 = 0.19$, min-max differences: $F_{(1, 48)} = 17.578$, $p_{Holm} < 0.01$, $\eta_p^2 = 0.268$) in that self-reported stressfulness was significantly higher during the TSST-VR than during the pTSST-VR (AUCi: $p_{Holm} < 0.001$, min-max differences: $p_{Holm} < 0.001$).

Physiological stress

Physiological stress parameters are illustrated in Figure 2, panel C and D. Exhaustive statistical parameters can be found in detail in the "Supplementary Material".

Cortisol

Comparing cortisol levels between groups over time resulted in a significant main effect of time ($F_{(1.8, 93.50)} = 4.378$, $p_{Holm} < 0.05$, $\eta_p^2 = 0.078$), but neither in a significant main effect of game-condition ($p_{Holm} > 0.999$), nor in a significant main effect of TSST-condition ($p_{Holm} = 0.964$, see Table S7). Furthermore, we did not observe any significant interaction effects for cortisol (all $p_{Holm} > 0.05$, see Table S7). Concerning the main effect of time, pairwise post-hoc comparisons did not survive corrections for multiple comparisons so that there was no difference in cortisol levels between any of the points in time (all $p_{Holm} > 0.05$). Concerning analyses of aggregated indices, AUCi and min-max differences rendered deviating results. AUCi of cortisol was significantly higher and positive in the TSST-VR compared to a negative AUCi of cortisol in the pTSST-VR ($p_{Holm} < 0.01$). In the omnibus-model, this was indicated by a main effect of TSST-condition ($F_{(1, 48)} = 7.26$, $p_{Holm} < 0.05$, $\eta_p^2 = 0.13$). In contrast, min-max differences did not render any statistically significant main or interaction effect (all $p_{Holm} > 0.05$).

Alpha-amylase

For sAA, none of the tested main or interaction effects reached statistical significance (all $p_{Holm} > 0.05$, see Table S8). Analyses of aggregated measures (AUCi and min-max differences) confirmed this pattern of null-findings (all $p_{Holm} > 0.05$).

Discussion

We investigated whether stress reactivity towards the OpenTSST VR (von Dawans et al., 2022) can be increased by prior exposure to a sensorimotor game in VR. In our sample, subjective measures (i.e., negative affect as assessed by means of the PANAS as well as self-reported stressfulness as assessed by means of the VAS) increased with exposure to the TSST-VR. In contrast, physiological measures (i.e., cortisol and sAA) did not respond well to the TSST-VR. This pattern did not differ between participants having played a sensorimotor game in VR and participants having completed a game in RL. Likewise, self-reported immersion was not affected by the gaming-manipulation.

Results of higher negative affect and self-reported stressfulness in the TSST-VR-group compared to the pTSST-VR-group were in line with our expectations: In accordance with existing literature, data confirm the OpenTSST VR to

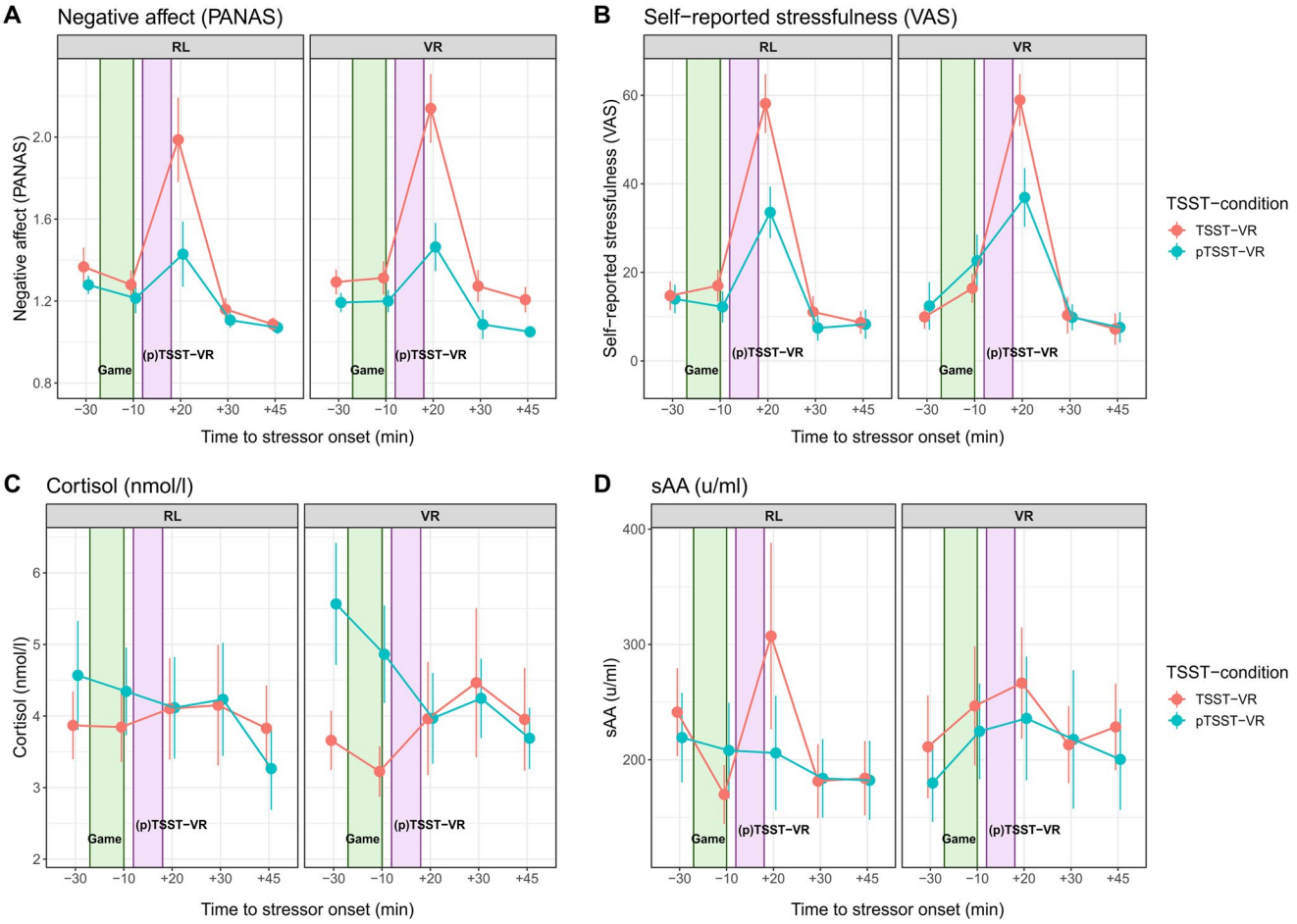

Figure 2. Repeated stress and affect measures.

Note. Depicted are repeated stress and affect measures over the course of the experiment, separate for all four experimental conditions (VR-game + TSST-VR, RL-game + TSST-VR, VR-game + pTSST-VR, RL-game + pTSST-VR), including the mean and standard error of the mean at all points in time (T1: -30 min, T2: -10 min, T3: +20 min, T4: +30 min, T5: +45 min relative to stressor onset). Duration of the game (either in VR or in RL) is illustrated in green. Duration of the TSST-VR or pTSST-VR is illustrated in violet. Each panel (A-D) is divided for game-conditions, with the RL-condition on the left, and the VR-condition on the right. TSST-conditions are illustrated using different colors: red lines illustrate trajectories within the TSST-VR-group, whereas blues line reflect dynamics for the pTSST-VR-group. A: negative affect, as assessed by means of the PANAS, B: self-reported stressfulness as given by means of the VAS, C: raw (untransformed) cortisol level (nmol/l), D: raw (untransformed) sAA activity (nmol/l).

successfully induce stress on a subjective dimension (e.g., Linnig et al., 2024; Shiban et al., 2016; Zimmer et al., 2019). In contrast to our hypothesis, cortisol and sAA did not respond well to the OpenTSST VR. Indeed, this pattern of results overlaps with a preprint by Linnig et al. (2024) validating the OpenTSST VR and its control condition in an independent project. Indeed, their data confirm that the OpenTSST VR might not trigger cortisol reactivity unambiguously. Meta-analytic evidence showed that TSST-VRs generally trigger cortisol reactivity. While effect sizes vary between individual studies, they consistently undercut in-person TSSTs with overall effects of medium size. Moreover, meta-analyses identified a publication bias toward larger effect sizes hinting at unpublished studies with small or no effects TSSTs (Helminen et al., 2019; van Dammen et al., 2022; but see Helminen et al., 2021). Consequently, the TSST-VR might not induce cortisol reactivity as reliably as the in-person TSST. The current results may be attributed to different factors.

Importantly, we included both sexes, and for women, we presumed intake of hormonal contraceptives. Following stress induction, women seem to respond less strong than men in terms of cortisol (in-person TSST: Liu et al., 2017; TSST-VR: Helminen et al., 2019; Santl et al., 2019) especially when taking oral contraceptives (in-person TSST: Liu et al., 2017; TSST-VR: Montero-López et al., 2018). However, including sex as a further factor (see Table S9 and Figure S1 in the "Supplementary Material") did not reveal a sex-specific pattern in additional analyses. Men and women seemed to respond comparably low towards the TSST-VR, so that sex may not explain dampened cortisol responsivity as reported for the whole sample. Moreover, two aspects limit the validity of sex-specific analyses. First, sex-specific analyses were underpowered for the current dataset since we did not plan to account for sex. Second, (if existent) sex-specific effects may not have been attributed to endocrine factors only (Jentsch et al., 2022).

Concerning cortisol, it is noteworthy that AUCi analyses as well as other explorative analyses using an alternative approach (no transformation of cortisol data) partly led to significant results concerning differences between TSST-VR and pTSST-VR over time. This suggests that the OpenTSST-VR might be capable to induce small cortisol reactivity principally. In our conventional approach, differences between the TSST-VR and the pTSST-VR might have not reached statistical significance since descriptively and by visual inspection of plotted data, cortisol

appeared to rise in the pTSST-VR group even though AUCi measures did not confirm this. Therefore, dynamics in the pTSST-VR group as a reference may have led to a certain blurring of effects. Indeed, the pTSST-VR condition might have included stressful elements. For instance, speaking in front of an empty room may have appeared bizarre and the math part might have challenged participants in requiring summation. As a result, control conditions other than the pTSST may be appropriate for VR. For instance, the friendly TSST (fTSST, Wiemers et al., 2013) contains an interacting panel that does not induce stress, but avoids the curiosity of speaking in front of an empty table. Additionally, the fTSST omits the math part. Another reason for stress reactivity during the fTSST may lie in the pure exposure to VR triggering some sort of *novelty effect* for participants being unfamiliar with VR.

Moreover, and unfortunately, baseline measures of cortisol and sAA were surprisingly high in the current dataset. This concerned the TSST-VR and the pTSST-VR group but was even increased in the latter (cf. Figure 2C). Thereby, initial measures provided an unfavorable baseline for further unbiased responses. Baseline values may reflect some sort of anticipatory stress. Excitement prior to testing is conceivable for participants as well as for student experimenters who engaged in data collection for the first time. To prevention initial stress reactivity, we scheduled testing sessions to the afternoon and instructed participants to avoid behaviors that could potentially activate stress systems (see "Method" section). Moreover, we randomly allocated participants to experimental conditions and established a double-blinded procedure in which experimenters were informed about a participant's condition shortly before the experiment. Lastly, beyond the intended manipulation, the game was considered an acclimatization period before further exposure. Still, our measures may not have sufficed to offset high initial levels in cortisol and systematic differences between the stress and the control group.

Furthermore, physiological stress parameters were only rarely responsive to our gaming manipulation. Initially, we assumed that completion of the VR-game may prepare processing of and receptivity for virtual input thereby leading to increased stress reactivity toward the TSST-VR. Alongside the observed null effects, also the IPQ did not reveal increased immersion in the group having played the VR-game. Interestingly, a recent study could improve neither perceived presence, nor stress reactivity increasing similarity between the laboratory environment and a TSST-VR (Zimmer et al., 2019). For the current data, different factors may come into play. First of all, the VR-game may not have sufficed to alter perceived immersion in our participants. In fact, the chosen VR-game was assumed to enable learning of *sensorimotor contingencies* as elaborated by Slater (2009). *Valid effective actions* were implemented in that participants used their controller to throw virtual balls into virtual cups. As virtual cups disappeared when being hit, the VR-game was further suggested to feature *valid sensorimotor actions*. Of note, Slater (2009) also emphasized the importance of movement in VR which we deemed to be implemented with the utilized VR-game. Still, these aspects may not have been appropriately captured by the VR-game. Furthermore, an exposure of 10 min may not have sufficed to learn *sensorimotor contingencies*. However, it is currently unknown which exposure-time triggers an immersion-effect. Lachlan and Krcmar (2011) who let participants play a video-game for different durations did not find effects on self-reported presence while another study by Melo et al. (2016) succeeded that way. Indeed, both procedures did not involve movement. Thus, it is unclear whether exposure to VR affects immersion, how long such an exposure should last and whether movement is moderator in such relations. An alternative explanation may consider the fact that *sensorimotor contingencies* noticeably differed between the VR-game and the TSST-VR and were not transferable. Indeed, participants completed a saliva sample and questionnaires between the VR-game and the TSST-VR. Thereby, participants did not set down the HMD, but they saw the SteamVR starting screen. Thus, it is conceivable that participants considered the VR-game and the TSST-VR as separate episodes that were not related, neither content-, nor modality-wise.

Overall, since manipulation of perceived immersion was not successful, one may ask how immersivity of TSST-VRs can be further increased assuming that it remains an adjusting screw for gathering larger effect sizes. Considering Helminen et al. (2019), it seems noteworthy that in their rationale, the current OpenTSST VR might have been evaluated as immersive already. In contrast, non-immersive TSST-VRs concerned procedures on 2D-screens. Vice versa, there is currently only one avenue to present TSST-VRs even more immersively: CAVEs. However, CAVEs are expensive (Creagh, 2003; Ronchi et al., 2019) and thus do conflict with benefits of TSST-VRs, namely resource-efficiency and accessibility. Still, more interactivity might improve immersivity of the OpenTSST VR. Even though not used in the current study, there is an eye-tracking option that enables judges to instruct participants to keep eye contact. In sum, future studies have to test immersivity as an adjusting screw for increased stress reactivity toward TSST-VRs and may clarify how immersivity can be enhanced targeting the TSST-VR procedure itself or contextual factors. As contextual factors, Schote et al. (2022), for instance, discussed conscious knowledge of an experimenter being present to add an external source of social evaluation. Finally, letting participants solve parallel cognitive tasks during exposure to VR might be effective as these could bind cognitive resources that might otherwise serve for questioning validity of the VR. Still, it is not clear how to realize cognitive tasks in parallel to demands of a TSST-VR itself.

The current study faced methodological constraints. As mentioned, data were acquired by students taking the position of experimenters. Thus, experimenters were less trained and probably varied more than normal (all students of the course had to engage in data collection) which may have impeded standardization. Still, we also see the strengths of this approach. On the one hand, the current study also served educational purposes. On the other hand, reduced standardization may be regarded a further sticking point to evaluate the true robustness of a stress induction procedure. Standardization may have also been impaired by the fact that we did not align the activation of avatars across experimenters. Indeed, there were no fixed rates on how often

experimenters should activate avatars to voice further statements so that this option might have been used with a certain variability. This leads to the question in how far a more frequent activation of avatars might have contributed to better stress reactivity in terms of physiological stress measures. Importantly, for the in-person TSST, it has been shown that stress reactivity was associated with the strength of the social evaluation as expressed by the panel (Vrshek-Schallhorn et al., 2018; Way & Taylor, 2010). Still, as already mentioned, we deemed activation of the virtual avatars as a tool promoting individuality and interactivity in that activation was utilized in response to the individual participant's behavior over time. Additionally, as mentioned earlier, we faced technical difficulties requiring recalibration of the VR for some participants. This implied a temporal delay and an interruption of the immersivity-manipulation (i.e., participants had to set down the HMD). However, a subsample without technical difficulties yielded the same results. Last but not least, we cannot rule out that the sample size was sufficiently large enough in order to disentangle hypothesized relations. Of note, our sample size was based on an a priori power calculation and we even surpassed the recommended number of participants. Still, there is an option that the included effect size (which was drawn from a meta-analysis) overestimated the true effect.

Conclusion

Adaptations of the TSST to VR open doors to new opportunities for stress research. However, it seems crucial to ensure TSST-VRs to trigger stress reactivity comparable to in-person applications of the TSST in order to be considered a genuine alternative. Moreover, against the background of various co-existing versions of TSST-VRs, research will benefit from a unified, openly accessible tool. The OpenTSST VR may satisfy this request but needs further validation and adaptation. As the current study suggests, the OpenTSST VR does not robustly lead to physiological stress reactivity, which may be partly attributed to inadequacy of the inbuilt control condition. While manipulation of perceived immersion independent of the OpenTSST VR did not lead to increased stress reactivity, upcoming work may prioritize optimization of the OpenTSST VR itself before targeting contextual variables. These may cover factors such as prior exposure to VR, similarity between VR and the real laboratory, as well as presence of experimenters as an external source of social-evaluative threat.

Acknowledgments

We acknowledge the support of Saskia Seel with respect to supporting us during the technical installation. We thank For Fun Labs for their contribution to Figure 1. Of note, the graphical abstract was created using Biorender.com.

Disclosure statement

The authors report there are no competing interests to declare.

Notes on contributors

Lena Sophie Pfeifer (M.Sc. in psychology and cognitive neuroscience, Ruhr University Bochum, 2020) is a PhD candidate at the Cognitive Psychology at Ruhr University Bochum. She is interested in different stress induction procedures (e.g., in virtual reality) and the effects of acute stress on functional laterality.

Carina Zoellner (M.Sc. in psychology and cognitive neuroscience, Ruhr University Bochum, 2019) is a PhD candidate at the Cognitive Psychology at Ruhr University Bochum. Her work focuses on virtual reality as a tool to investigate the interplay of semantic and episodic information during the retrieval of a past episodes and according neural correlates.

Oliver T. Wolf (diploma in psychology, Trier University, 1995; PhD Trier University, 1997, habilitation, Heinrich Heine University Düsseldorf, 2001) is a full professor of Cognitive Psychology at Ruhr University Bochum. He investigates the influence of stress hormones on cognition and emotion.

Gregor Domes (diploma in psychology, Eberhard Karls University Tübingen, 1999; PhD at the Eberhard Karls University Tübingen, 2003, habilitation, Albert Ludwigs University Freiburg, 2012) is a full professor of Biological and Clinical Psychology at the Trier University. He investigates the endocrine and neural underpinnings of social cognition and behavior.

Christian J. Merz (diploma in psychology, Justus Liebig University Giessen, 2007; PhD Ruhr University Bochum, 2011; habilitation, Ruhr University Bochum, 2017) is an associate professor of Cognitive Psychology at Ruhr University Bochum. His research program focuses on the influence of stress and sex hormones on (emotional) learning and memory processes.

Funding

The contribution of LSP and OTW was supported by the DFG within project B4 of the Collaborative Research Center (SFB) 874 "Integration and Representation of Sensory Processes" [project number 122679504] as well as within the Open Research Area (ORA; WO 733/17-1; project number: 396877307). CZ and OTW were supported by the DFG within the project P4 of the Research Group (FOR) 2812 "Constructing scenarios of the past: A new framework in episodic memory" [project number 397530566]. GD was supported by the Trier University Research Priority Program "Psychobiology of Stress" funded by the State Rhineland Palatinate.

ORCID

Lena Sophie Pfeifer http://orcid.org/0000-0001-8803-6515
Carina Zoellner http://orcid.org/0000-0002-0775-2147
Oliver T. Wolf http://orcid.org/0000-0002-9320-2124
Gregor Domes http://orcid.org/0000-0001-5908-4374
Christian J. Merz http://orcid.org/0000-0001-5679-6595

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*, 1–14. https://doi.org/10.1016/j.ynstr.2016.11.001

Andri, S. et al. (2022). *DescTools: Tools for descriptive statistics*. Retrieved from https://cran.r-project.org/package=DescTools

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of Effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Breyer, B., & Bluemke, M. (2016). *Deutsche version der positive and negative affect schedule PANAS (GESIS Panel)*. https://doi.org/10.6102/ZIS242

Creagh, H. (2003). Cave automatic virtual environment. In *Proceedings: Electrical Insulation Conference and Electrical Manufacturing and Coil*

Winding Technology Conference (Cat. No.03CH37480) (pp. 499–504). IEEE. https://doi.org/10.1109/EICEMC.2003.1247937

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*(3), 355–391. https://doi.org/10.1037/0033-2909.130.3.355

Doane, L. D., Kremen, W. S., Eaves, L. J., Eisen, S. A., Hauger, R., Hellhammer, D., Levine, S., Lupien, S., Lyons, M. J., Mendoza, S., Prom-Wormley, E., Xian, H., York, T. P., Franz, C. E., & Jacobson, K. C. (2010). Associations between jet lag and cortisol diurnal rhythms after domestic travel. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *29*(2), 117–123. https://doi.org/10.1037/a0017865

Elkin, L. A., Kay, M., Higgins, J. J., & Wobbrock, J. O. (2021). *An aligned rank transform procedure for multifactor contrast tests*. ACM Press. https://doi.org/10.1145/3472749.3474784

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Ferrara, A., Rapuano, M., Ruotolo, F., Ruggiero, G., & Iachini, T. (2021). Being there in virtual reality: the sense of presence is associated with the ability to imagine body movements. *Psychology-Based Technologies PSYCHOBIT*.

Goodman, W. K., Janson, J., & Wolf, J. M. (2017). Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology*, *80*, 26–35. https://doi.org/10.1016/j.psyneuen.2017.02.030

Grosser, L., Knayfati, S., Yates, C., Dorrian, J., & Banks, S. (2022). Cortisol and shiftwork: A scoping review. *Sleep Medicine Reviews*, *64*, 101581. https://doi.org/10.1016/j.smrv.2021.101581

Helminen, E. C., Morton, M. L., Wang, Q., & Felver, J. C. (2019). A meta-analysis of cortisol reactivity to the Trier Social Stress Test in virtual environments. *Psychoneuroendocrinology*, *110*, 104437. https://doi.org/10.1016/j.psyneuen.2019.104437

Helminen, E. C., Morton, M. L., Wang, Q., & Felver, J. C. (2021). Stress reactivity to the trier social stress test in traditional and virtual environments: A meta-analytic comparison. *Psychosomatic Medicine*, *83*(3), 200–211. https://doi.org/10.1097/PSY.0000000000000918

Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., & Wolf, O. T. (2009). Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. *Psychoneuroendocrinology*, *34*(7), 1075–1086. https://doi.org/10.1016/j.psyneuen.2009.02.008

Hoogerwerf, M. D., Veldhuizen, I. J. T., Kort, W. L. A. M. D., Frings-Dresen, M. H. W., & Sluiter, J. K. (2015). Factors associated with psychological and physiological stress reactions to blood donation: A systematic review of the literature. *Blood Transfusion = Trasfusione Del Sangue*, *13*(3), 354–362. https://doi.org/10.2450/2015.0139-14

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Jentsch, V. L., Pötzl, L., Wolf, O. T., & Merz, C. J. (2022). Hormonal contraceptive usage influences stress hormone effects on cognition and emotion. *Frontiers in Neuroendocrinology*, *67*, 101012. https://doi.org/10.1016/j.yfrne.2022.101012

Kassambara, A. (2022a). *ggpubr: 'ggplot2' based publication ready plots*. Retrieved from https://CRAN.R-project.org/package=ggpubr

Kassambara, A. (2022b). *rstatix: Pipe-friendly framework for basic statistical tests*. Retrieved from https://CRAN.R-project.org/package=rstatix

Kirschbaum, C., Pirke, K. M., & 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

Krohne, H. W., Egloff, B., Kohlmann, C. W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS) [Investigations with a German version of the Positive and Negative Affect Schedule (PANAS)]. *Diagnostica*, *42*(2), 139–156. https://doi.org/10.1037/t49650-000

Kudielka, B. M., Hellhammer, D. H., & Kirschbaum, C. (Eds.) (2007). *Social neuroscience: Integrating biological and psychological explanations of social behavior*. In *Ten years of research with the trier social stress test–revisited* (pp. 56–83). The Guilford Press.

Kudielka, B. M., Hellhammer, D. H., & Wüst, S. (2009). Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. *Psychoneuroendocrinology*, *34*(1), 2–18. https://doi.org/10.1016/j.psyneuen.2008.10.004

Labuschagne, I., Grace, C., Rendell, P., Terrett, G., & Heinrichs, M. (2019). An introductory guide to conducting the Trier Social Stress Test. *Neuroscience and Biobehavioral Reviews*, *107*, 686–695. https://doi.org/10.1016/j.neubiorev.2019.09.032

Lachlan, K., & Krcmar, M. (2011). Experiencing presence in video games: The role of presence tendencies, game experience, gender, and time spent in play. *Communication Research Reports*, *28*(1), 27–31. https://doi.org/10.1080/08824096.2010.518924

Linnig, K., Seel, S., von Dawans, B., Zielasko, D., Standard, W., Weyers, B., & Domes, G. (2024). *Open TSST VR: Psychobiological reactions to an open version of the trier social stress test in virtual reality*. https://doi.org/10.31234/osf.io/fpe2h

Liu, J. J. W., Ein, N., Peck, K., Huang, V., Pruessner, J. C., & Vickers, K. (2017). Sex differences in salivary cortisol reactivity to the Trier Social Stress Test (TSST): A meta-analysis. *Psychoneuroendocrinology*, *82*, 26–37. https://doi.org/10.1016/j.psyneuen.2017.04.007

Lorentz, K., Gütschow, B., & Renner, F. (1999). Evaluation of a direct alpha-amylase assay using 2-chloro-4-nitrophenyl-alpha-D-maltotrioside. *Clinical Chemistry and Laboratory Medicine*, *37*(11-12), 1053–1062. https://doi.org/10.1515/CCLM.1999.154

Mangiafico, S. S. (2023). *rcompanion: Functions to Support Extension Education Program Evaluation*. Retrieved from https://CRAN.R-project.org/package=rcompanion/

Melo, M., Sampaio, S., Barbosa, L., Vasconcelos-Raposo, J., & Bessa, M. (2016). The impact of different exposure times to 360° video experience on the sense of presence. In *2016 23° Encontro Português de Computação Gráfica e Interação (EPCGI)* (pp. 1–5). IEEE. https://doi.org/10.1109/EPCGI.2016.7851192

Miller, R., Wojtyniak, J. G., Weckesser, L. J., Alexander, N. C., Engert, V., & Lehr, T. (2018). How to disentangle psychobiological stress reactivity and recovery: A comparison of model-based and non-compartmental analyses of cortisol concentrations. *Psychoneuroendocrinology*, *90*, 194–210. https://doi.org/10.1016/j.psyneuen.2017.12.019

Montero-López, E., Santos-Ruiz, A., García-Ríos, M. C., Rodríguez-Blázquez, M., Rogers, H. L., & Peralta-Ramírez, M. I. (2018). The relationship between the menstrual cycle and cortisol secretion: Daily and stress-invoked cortisol patterns. *International Journal of Psychophysiology: official Journal of the International Organization of Psychophysiology*, *131*, 67–72. https://doi.org/10.1016/j.ijpsycho.2018.03.021

Nater, U. M., La Marca, R., Florin, L., Moses, A., Langhans, W., Koller, M. M., & Ehlert, U. (2006). Stress-induced changes in human salivary alpha-amylase activity – Associations with adrenergic activity. *Psychoneuroendocrinology*, *31*(1), 49–58. https://doi.org/10.1016/j.psyneuen.2005.05.010

Nater, U. M., & Rohleder, N. (2009). Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: Current state of research. *Psychoneuroendocrinology*, *34*(4), 486–496. https://doi.org/10.1016/j.psyneuen.2009.01.014

Nater, U. M., Rohleder, N., Gaab, J., Berger, S., Jud, A., Kirschbaum, C., & Ehlert, U. (2005). Human salivary alpha-amylase reactivity in a psychosocial stress paradigm. *International Journal of Psychophysiology: official Journal of the International Organization of Psychophysiology*, *55*(3), 333–342. https://doi.org/10.1016/j.ijpsycho.2004.09.009

Phillips, A. C., Carroll, D., Burns, V. E., & Drayson, M. (2005). Neuroticism, cortisol reactivity, and antibody response to vaccination. *Psychophysiology*, *42*(2), 232–238. https://doi.org/10.1111/j.1469-8986.2005.00281.x

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, *28*(7), 916–931. https://doi.org/10.1016/s0306-4530(02)00108-7

Regenbrecht, H., & Schubert, T. (2002). Real and illusory interactions enhance presence in virtual environments. *Presence*, *11*(4), 425–434. https://doi.org/10.1162/105474602760204318

Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research*. Retrieved from https://CRAN.R-project.org/package=psych

Rezai, M., Fullwood, C., Hird, B., Chawla, M., Tetlow, L., Banerjee, I., & Patel, L. (2022). Cortisol levels during acute illnesses in children and adolescents: A systematic review. *JAMA Network Open*, *5*(6), e2217812. https://doi.org/10.1001/jamanetworkopen.2022.17812

Rohleder, N., Nater, U. M., Wolf, J. M., Ehlert, U., & Kirschbaum, C. (2004). Psychosocial stress-induced activation of salivary alpha-amylase: An indicator of sympathetic activity? *Annals of the New York Academy of Sciences*, *1032*(1), 258–263. https://doi.org/10.1196/annals.1314.033

Ronchi, E., Mayorga, D., Lovreglio, R., Wahlqvist, J., & Nilsson, D. (2019). Mobile-powered head-mounted displays versus cave automatic virtual environment experiments for evacuation research. *Computer Animation and Virtual Worlds*, *30*(6), e1873. https://doi.org/10.1002/cav.1873

RStudio Team. (2021). *RStudio: Integrated development environment for R*. Retrieved from http://www.rstudio.com/

Santl, J., Shiban, Y., Plab, A., Wüst, S., Kudielka, B. M., & Mühlberger, A. (2019). Gender Differences in Stress Responses during a Virtual Reality Trier Social Stress Test. *International Journal of Virtual Reality*, *19*(2), 2–15. https://doi.org/10.20870/IJVR.2019.19.2.2912

Schote, A. B., Dietrich, K., Linden, A. E., Dzionsko, I., Molano Moreno, L. D. L. A., Winnikes, U., Zimmer, P., Domes, G., & Meyer, J. (2022). Real sweating in a virtual stress environment: Investigation of the stress reactivity in people with primary focal hyperhidrosis. *PloS One*, *17*(8), e0272247. https://doi.org/10.1371/journal.pone.0272247

Schubert, T. W. (2003). The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Journal of Media Psychology*, *15*(2), 69–71. https://doi.org/10.1026//1617-6383.15.2.69

Shiban, Y., Diemer, J., Brandl, S., Zack, R., Mühlberger, A., & Wüst, S. (2016). Trier Social Stress Test in vivo and in virtual reality: Dissociation of response domains. *International Journal of Psychophysiology: official Journal of the International Organization of Psychophysiology*, *110*, 47–55. https://doi.org/10.1016/j.ijpsycho.2016.10.008

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Usoh, M., & Steed, A. (1995). Taking steps: The influence of a walking metaphor on presence in virtual reality. *ACM Transactions on Computer Human Interaction*, *2*, 201–219. https://doi.org/10.1145/210079.210084

Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., & Brooks, F. P. (1999). Walking > walking-in-place > flying, in virtual environments. In W. Waggenspack (Ed.), *Proceedings of the 26th annual conference on Computer graphics and interactive techniques – SIGGRAPH '99* (pp. 359–364). ACM Press. https://doi.org/10.1145/311535.311589

van Dammen, L., Finseth, T. T., McCurdy, B. H., Barnett, N. P., Conrady, R. A., Leach, A. G., Deick, A. F., Van Steenis, A. L., Gardner, R., Smith, B. L., Kay, A., & Shirtcliff, E. A. (2022). Evoking stress reactivity in virtual reality: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *138*, 104709. https://doi.org/10.1016/j.neubiorev.2022.104709

von Dawans, B., Standard, W., Zielasko, D., Seel, S., Weyers, B., & Domes, G. (2022). The Open TSST VR. An open tool for inducing psychosocial stress in virtual reality. Retrieved from http://www.Uni-Trier.De/index.Php?Id=88863

Vrshek-Schallhorn, S., Avery, B. M., Ditcheva, M., & Sapuram, V. R. (2018). The cortisol reactivity threshold model: Direction of trait rumination and cortisol reactivity association varies with stressor severity. *Psychoneuroendocrinology*, *92*, 113–122. https://doi.org/10.1016/j.psyneuen.2017.11.002

Way, B. M., & Taylor, S. E. (2010). The serotonin transporter promoter polymorphism is associated with cortisol response to psychosocial stress. *Biological Psychiatry*, *67*(5), 487–492. https://doi.org/10.1016/j.biopsych.2009.10.021

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H., & Girlich, M. (2022). *tidyr: Tidy Messy data*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wiemers, U. S., Sauvage, M. M., Schoofs, D., Hamacher-Dang, T. C., & Wolf, O. T. (2013). What we remember from a stressful episode. *Psychoneuroendocrinology*, *38*(10), 2268–2277. https://doi.org/10.1016/j.psyneuen.2013.04.015

Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., & Domes, G. (2019). Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology*, *101*, 186–192. https://doi.org/10.1016/j.psyneuen.2018.11.010

Zimmer, P., Wu, C. C., & Domes, G. (2019). Same same but different? Replicating the real surroundings in a virtual trier social stress test (TSST-VR) does not enhance presence or the psychophysiological stress response. *Physiology & Behavior*, *212*(Article 112690), 112690. https://doi.org/10.1016/j.physbeh.2019.112690