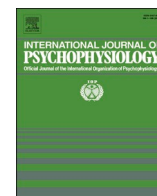


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Multi-modal responses to the Virtual Reality Trier Social Stress Test: A comparison with standard interpersonal and control conditions

M.A. Fallon^a, M.M.E. Riem^{b,c,d}, L.E. Kunst^b, W.J. Kop^b, N. Kupper^{b,*}^a Business School, University of Mannheim, Mannheim, Germany^b Center of Research on Psychological & Somatic Disorders, Department of Medical and Clinical Psychology, Tilburg University, Tilburg, the Netherlands^c Behavioral Science Institute, Radboud University, the Netherlands^d Clinical Child & Family Studies, Faculty of Behavioral and Movement Sciences, Vrije Universiteit, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:

Trier Social Stress Test
 Virtual reality
 Psychosocial stress
 Cardiovascular responses
 Heart rate variability

ABSTRACT

The Trier Social Stress Test (TSST) is a reliable social-evaluative stressor. To overcome limitations of the *in vivo* TSST, a standardized virtual reality TSST (VR-TSST) was developed. The present study compares the emotional (anxiety) and physiological (heart period and variability) response to a VR-TSST with an *in vivo* TSST and a control condition. Participants took part in either an *in vivo* TSST ($N = 106$, 64% female), VR-TSST ($N = 52$, 100% female), or a control TSST ($N = 20$, 40% female). Mixed linear modeling examined response profile differences related to TSST type. While there was an equivalent anxiety response to the *in vivo* TSST as the VR-TSST, we found a smaller heart period and heart rate variability response in VR-TSST compared to the *in vivo* TSST, especially in response to the math part of the test. The present findings demonstrate that social evaluative stress can be successfully induced in a VR setting, producing similar emotional and slightly attenuated cardiovascular responses.

1. Introduction

The association between psychological distress and disease, particularly cardiovascular disease, is well-established in research (Cohen et al., 2007). The mechanisms underlying this association have been studied in laboratory experiments, illustrating that acute lab stressors provoke disease-relevant physiological responses. One of the most reliable and widely used laboratory tests for studying the physiological stress response is the Trier Social Stress Test (TSST) (Dickerson and Kemeny, 2004). The TSST is administered in an experimental setting, which involves a socially-evaluated public speaking and a mental arithmetic task, and is conducted face-to-face with trained evaluators. Cognitive effort, the setting, and the social evaluation aspect all contribute to the observed physiological stress response (Kirschbaum et al., 1993). The TSST reliably induces neuroendocrine, cardiovascular, and immune responses (Allen et al., 2014). The observed physiological response evoked by the TSST makes it a valuable tool for studying biological mechanisms relevant to the association between psychosocial factors and adverse disease progression.

Conducting the TSST in a laboratory setting (*in vivo*) has several

limitations that can interfere with the use of this procedure in a broad range of clinically relevant settings. First, variations in the administration of the TSST show differential impacts on physiological responses. Variations in the attentiveness of the evaluators and the extent to which they are critical of the participant can affect the observed physiological responses (Goodman et al., 2017). Limiting and controlling for variations in facial expressions and body gestures of evaluators can, however, be difficult in a face-to-face experimental setting and across different laboratories. Second, the stationary nature of the TSST may be limiting its capability to measure stress responses outside a laboratory setting. For example, participants and evaluators need to be in the same location, which means it cannot easily be administered in a wide variety of research environments (e.g. functional magnetic resonance imaging scanner or field applications).

To overcome the limitations of the *in vivo* TSST, researchers have utilized a virtual reality TSST (VR-TSST). There has been limited research, however, comparing the emotional and physiological responses to a VR-TSST with an *in vivo* TSST. Illustrating that a VR-TSST can produce reliable emotional and physiological responses comparable to an *in vivo* TSST would lend credibility to the utilization of a VR-

* Corresponding author at: Center of Research on Psychological & Somatic Disorders, Department of Medical and Clinical Psychology, Tilburg University, the Netherlands.

E-mail address: H.M.Kupper@tilburguniversity.edu (N. Kupper).

<https://doi.org/10.1016/j.ijpsycho.2021.01.010>

Received 2 July 2020; Received in revised form 12 January 2021; Accepted 12 January 2021

Available online 14 January 2021

0167-8760/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

TSST in a broad array of clinical and research settings. Such findings could also help researchers to better understand the mechanisms linking psychosocial stressors to disease by studying these processes outside of a laboratory setting. Moreover, the VR-TSST could eliminate variations in attentiveness and criticalness of evaluators' through a more controlled research environment.

It may not be immediately apparent why a VR-TSST would produce a physiological response comparable to an *in vivo* TSST. Because participants are, in a way, observing the psychosocial stressor via the VR device, rather than directly experiencing it face-to-face, they might feel less threatened by it. However, research suggests that social engagement in virtual reality can indeed be realistic and comparable to real-life interactions (Grinberg et al., 2014; Guadagno et al., 2007; Knowles et al., 2017). Moreover, realistic social interactions in virtual reality increase feelings of immersion, which is associated with users' motivation and engagement of the virtual reality space (Schultze, 2010). Thus, the social interactions in a VR-TSST could suffice to activate representations of reality and sufficiently produce a physiological stress response comparable to an *in vivo* TSST.

Prior research has attempted to validate versions of a VR-TSST (Fallon et al., 2016; Jonsson et al., 2010; Kotlyar et al., 2008; Montero-Lopez et al., 2016; Ruiz et al., 2010). However, only few studies directly compare cardiovascular responses of an *in vivo* TSST to VR-TSST. Studies that have made a direct comparison report similar to somewhat smaller increases in heart rate (Shiban et al., 2016; Zimmer et al., 2019) and heart rate variability in response to a speech task (Kothgassner et al., 2016). However, the virtual reality condition in these studies used head-mounted technology, which can cause participants to experience nausea and simulation sickness (Pan and Hamilton, 2018), conflating the stress of vertigo with the psychosocial stress evoked by the TSST. Moreover, head-mounted virtual reality systems are expensive and are not feasible to use in unique research environments, such as magnetic resonance imaging (MRI) scanners or in field applications (Wilson and Soranzo, 2015). In the present study, we therefore investigate the effects of a VR-TSST that requires no confederates, using the Second Life platform, in comparison with an *in vivo* TSST and a control TSST on cardiovascular activity (heart period and heart rate variability) and negative emotion.

The current study uses an inexpensive and widely available online virtual world technology presented on a computer screen as opposed to a head-mounted display. To our knowledge, this study is the first to compare stress induced by a virtual TSST without head-mounted technology with stress levels induced by an *in vivo* TSST. If it were possible to produce similar cardiovascular responses to a real-world TSST, it would have several practical and experimental advantages (e.g. lower costs, more standardization, and ability to evaluate stress responses outside a laboratory). We hypothesize that the VR-TSST, using the Second Life platform on a computer screen, will evoke changes in emotional and cardiovascular reactivity similar to an *in vivo* TSST (H1). Moreover, we expect that responses from the *in vivo* TSST will be significantly more pronounced than responses produced from a control TSST without psychosocial challenges (H2).

2. Methods

2.1. Participants

The present study involves the analysis of participants in two investigations: A study on the effects of oxytocin on stress reactivity (Study 1) and a second project (PHEMORE study) examining individual differences in stress reactivity (Study 2 and Study 3).

Study 1 included 52 female undergraduate students (Mean age = 19.9 ± 1.8) participated in the Virtual TSST study. The majority (76.9%) did not know the Second Life platform, some participants had heard of it or seen it (21.1%), and only one participant had an account. The focus of the larger project was to examine intranasal oxytocin effects on stress

reactivity and all participants were female (Riem et al., 2020). All participants selected for the present analyses received a placebo nasal spray (double-blind). Participants were screened for drug or alcohol abuse, nasal problems, use of prescribed medication (except contraceptive medication), psychiatric and neurological disorders, cardiovascular diseases, and hypertension. Furthermore, participants who were pregnant, breastfed or had children were excluded from this study. The study was approved by the Brabant Medical Ethics Committee (NL60593.028.17).

Data for the *in-vivo* TSST subset (Study 2) were retrieved from the larger PHEMORE (Physiological and EMotional Reactivity) study (Kupper et al., 2020), which examined individual differences in reactivity to mental stress among young adults. Data collection for this larger study went on from January 2011 until June 2016 and was described earlier in more detail (Kupper et al., 2020). Other individual differences oriented papers published on a selection of the PHEMORE dataset either are published (Duijndam et al., 2020), or are in the process of being written. From the regular TSST dataset from 2015 and 2016 (closest in time to the VR data collection), we drew a sample of 106 young adults (36% male, age $M = 20.5 \pm 2.8$), taking into consideration that the task order in which they completed the experiment was similar as the VR version (*i.e.* speech first, then math). To test the first hypothesis, we selected the women from this sample ($n = 68$). For the second hypothesis, the full sample was used. Post-hoc power analysis for hypothesis 1 (most limited sample) showed that with an alpha of 0.05, we had sufficient statistical power (0.95) to prove a medium effect size ($f = 0.15$ /Cohen's $d = 0.40$).

In 2016, 20 participants took part in a 'control TSST' (Study 3), as part of the PHEMORE study, in which active stressor elements were removed (aged $M = 21.2$, $SD = 1.9$, 40% female). Age did not differ across studies, $F(2, 137) = 1.86$, $p = .159$. This sample was only used to test hypothesis 2. The Institutional Ethics Review Board approved the PHEMORE study protocol (Study 2 and 3) and its amendments (protocol number: EC-2011.01a). All participants gave informed consent before participating and were debriefed afterwards.

2.2. Procedure

2.2.1. Study 1

2.2.1.1. Virtual TSST. The present data were collected in the GO-Lab of Tilburg University, as part of a larger study on stress reactivity and oxytocin (Riem et al., 2020). Participants were instructed to refrain from smoking and coffee consumption on the day of the lab session and from alcoholic beverages during the 24 h before testing. The VR-TSST protocol was highly similar to the protocol of the *in vivo* and control TSST (see Fig. 1 and Appendix A). After signing informed consent, ECG electrodes were attached and participants completed a 5-min rest baseline measure, while watching a picture depicting a nature scene, after which participants completed self-report anxiety measures (baseline). Subsequently, the experimenter read out the instructions for the TSST. The original protocol of the TSST was adapted such that participants were to remain seated throughout the entire procedure, because a standing position or changes in posture may cause fluctuations in blood pressure (Olufsen et al., 2005). The VR-TSST was conducted using the Second Life platform. Second Life is an online virtual world. Within Second Life, individuals can interact with one another using virtual representations of themselves (called avatars) through audio and chat functions. Although this platform is public, the area that was used for the VR-TSST was private and could not be accessed by anyone other than those permitted by the principal investigator. Participants were instructed to imagine applying for an internship position through the Second Life platform. They were asked to prepare a 5-minute speech to convince two professors that they would be the ideal candidate for the position. The participant and the two professors were represented as

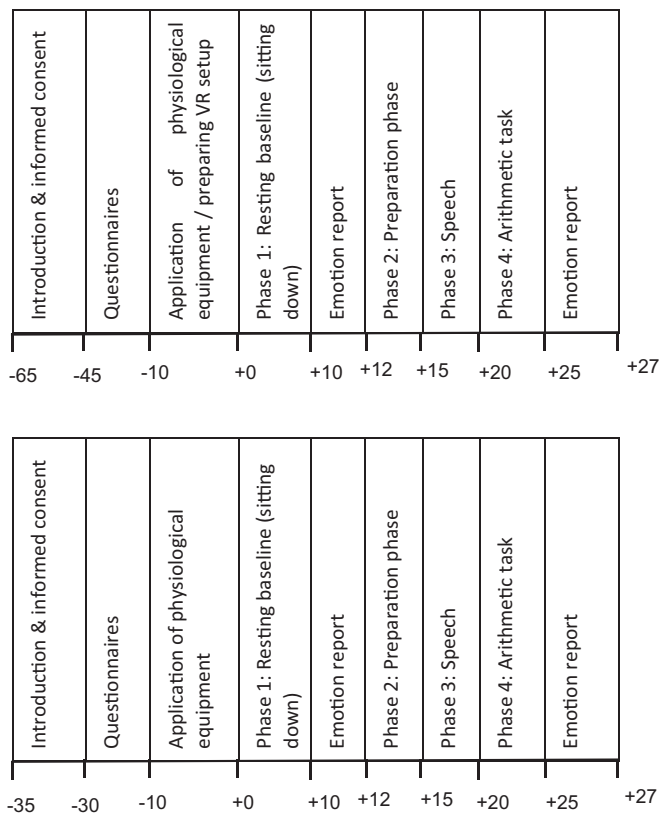


Fig. 1. VR-TSST and *in vivo* TSST test procedure.

Note: Experimental procedure indicating the subsequent experimental phases and their timing for respectively the VR (top panel) and the *in vivo* TSST (lower panel).

their own avatar in Second Life. After the speech, an additional math task would provide information about the applicant's working memory capacity. A 5-minute preparatory period started after the instructions, in which the experimenter retreated to the observation room. After 5 min, the experimenter showed the participant the Second Life environment: the TSST took place in a large auditorium with a virtual stage (See Appendix A; (Fallon et al., 2016)). Participants stood on the lower part of the stage and looked at the two front seats, in which one male and one female professor avatars were seated. The visual settings were zoomed in, so that participants only saw the professors. The experimenter told the participant that (s)he would briefly contact the professors, to verify that they had logged in successfully, left the lab room and announced through a microphone that the professors were ready and would be in contact in a minute.

The experimenter controlled the two professor avatars in Second Life. Second Life allows pre-recorded audio messages to be uploaded and then played by using the Sounds function. We recorded 36 Dutch and 36 English (for international students) messages that followed TSST protocols as described by Kupper et al. (Kupper et al., 2020) and used in Study 2 (based on PHEMORE). The messages were recorded such that the male and female professors talked in equal proportions. The first recordings included a brief introduction (Female: 'Hi, can you hear me?', 'Ok, we will begin the task shortly'). The male professor then instructed participants to start their speech. The following prompts were played if participants were silent for 3 s: female: 'You still have some time, please continue', male: 'You still have time, go on', male: 'Can you tell us something about your strengths?', female: 'How would other students describe your social skills?', and male: 'Can you tell me something about your weaknesses?'. In line with Fallon et al. (Fallon et al., 2016), the professor avatars used the gestures 'bored' twice and 'shrug' once, at 1, 3 and 4 min into the speech respectively.

After 5 min, the professors gave the instructions for the math task. The math task entailed subtracting 13 from a number, and then repeatedly subtracting 13 from the remainder. Upon each mistake a new starting number was given (e.g., 'That's incorrect, please start again, and this time start from 1072'). Additionally, the following prompts were used once per participant: 'At this point, you're making more errors compared with other participants. Try to be more accurate' and 'You are being a little slow compared to the other participants. Please try to speed up your answers'. After 5 min of performing the math task, the male avatar indicated 'Please stop, your time is up. You can tell the experimenter now that you are finished (instructed to raise hand)'. The remaining messages were recorded to have a variety of options in case participants behaved unexpectedly (e.g., 'Are you OK to continue?', 'I cannot comment on that', 'Yes, we can hear you').

After the math task, the experimenter returned to the laboratory room and administered the second self-report anxiety measure. Participants underwent a debriefing procedure at the end of the protocol. The experimenter asked a series of increasingly suggestive questions to uncover whether the participant believed that she was talking to two professors (i.e., What was your impression of the two professors?).

2.2.2. Study 2

2.2.2.1. In vivo TSST. Participants were instructed to refrain from smoking and coffee consumption for 2 h before testing as well as not to ingest more than three alcoholic beverages during the 24 h before testing. After providing informed consent, participants were fitted with the cardiovascular measurement equipment at the GO-Lab. After a 10-min resting period, during which we recorded a physiological baseline, participants took part in a 5-min cognitive task not related to the present analyses and a recovery period (5 min). The stress-inducing part of the protocol then started using the Trier Social Stress Test (TSST), followed by a 5-min recovery period. Participants filled out a second questionnaire at the end of the protocol. The present paper reports on the results pertaining to the 10-min resting phase and the responses to the TSST.

We adapted the original protocol of the TSST in two ways. First, similar to Study 1, we asked participants to remain seated throughout the entire procedure (Olufsen et al., 2005). Second, instead of a job interview, we asked participants to prepare (three-minute preparation period) and give a speech on their own positive and negative social skills (five minutes), in front of a two-person audience. Previous research has shown that the current procedure induces a significant cardiovascular stress response (Kupper et al., 2013). During the arithmetic task, participants were asked to serially subtract a one-digit (e.g., 7) or two digit (e.g., 13 or 19) numbers from four digit numbers verbally in the presence of a socially evaluative audience. Comments were scripted and are presented in Supplement 1.

2.2.3. Study 3

2.2.3.1. Control TSST. The control TSST was designed to be as close to the original TSST as possible, while removing the key stress-inducing elements, similar to the procedure described by Het et al. (Het et al., 2009). After a 10-minute resting period, the participant was asked to give a 5-minute speech about a movie, novel, a recent holiday trip, or what they did during the weekend. The participants were informed that there would be a 3-minute preparation period during which they should think about the topic of the speech. After 3 min, the experimenter entered the room and asked the participant to start their speech. The experimenter stayed in the room, listening and nodding empathically. If a participant stopped talking, the experimenter first asked whether he/she could tell some more, and if not, asked a question. After 5 min, the experimenter asked the participant to stop talking and to start adding up the number 5 starting at 0. This second task also lasted 5 min. The

experimenter wrote down the highest number reached. When the math task was finished, participants were asked to sit and rest (recovery period) for 5 min. The control TSST was performed in the same lab room at GO-Lab as the *in-vivo* TSST and the VR-TSST, but all ‘stressing’ elements of the TSST (committee of evaluators, performance pressure) were removed. This procedure was expected to eliminate the main effective factors of the TSST, namely the social evaluative threat and the uncontrollability, consistent with the theory proposed by Dickerson and Kemeny (2004).

2.3. Materials and instruments

2.3.1. Self-reported anxiety

Anxiety was measured after the resting baseline and right after the TSST math task in all three studies. In the virtual TSST study (Study 1), we used the Spielberger Trait-State Anxiety Inventory, State version (STAI). The STAI includes 6 items that are scored on a 4-point Likert scale (Marteau and Bekker, 1992). Participants were asked to complete the STAI directly after the math task, and were asked to indicate how they were feeling at that moment. We calculated a total score for STAI-S.

In Study 2 and 3, anxiety was measured using four 7-point Likert scale items on anxiety. Participants were asked to indicate to what extent they felt these emotions during the preceding task (after resting baseline, and after the stress battery).

To enable comparison between the outcome measures of Study 1 and Studies 2 and 3, we selected four items of the STAI (Study 1) that matched the anxiety items of Study 2 and 3. In Study 2 and 3 the items were ‘I feel at ease’ (reversed), ‘I am tense’, ‘I feel anxious’, ‘I am stressed’. Therefore, we selected the following items from the STAI in Study 1: ‘I feel calm’ (reversed), ‘I am tense’, ‘I feel upset’ and ‘I am relaxed’ (reversed). The internal consistency of the four-item STAI in Study 1 was $\alpha = 0.73$ at the resting baseline as well as after the math task. The internal consistency of the derived four items anxiety scale in Study 2 and 3 was $\alpha = 0.82$ at baseline and $\alpha = 0.88$ after the math task.

2.4. Electrocardiogram

In Study 1, the VR-TSST, heart period and heart rate variability were derived from continuous ECG recordings made with the ECG100C module and the Biopac MP150 system, and three hydrogel ECG electrodes. Data were recorded at a sampling frequency of 2000 Hz. Data processing was conducted in AcqKnowledge, version 4.4. Human ECG complex boundaries were identified automatically and artifacts and missed QRS peaks were identified and corrected manually. We calculated period averages for heart period (IBI), beats per minute (BPM), and the average root mean square of successive differences (RMSSD), a measure of cardiac parasympathetic activation, for each experiment phase.

In Study 2 and 3, the PHEMORE (*in vivo* and control TSST studies), the Vrije Universiteit Ambulatory Monitoring System (VU-AMS 4.6; Vrije Universiteit Amsterdam, the Netherlands) was used to record a continuous electrocardiogram (ECG) and impedance cardiogram (ICG) at a frequency of 1000 Hz (Z0 at 250 Hz) (De Geus et al., 1995). Seven non-woven, liquid gel AgCl electrodes (Kendall, Medcat, the Netherlands) were used. The event button on the device was used to indicate start and end times of the phases of the experimental protocol and was operated by the test leader based on a stopwatch timing. VU-AMS software was used to automatically detect all R-peaks in the ECG, and all R-peak markers were visually checked and adjusted manually when necessary. The signal was visually checked for artifacts (e.g., premature atrial or ventricular contractions), which were removed prior to scoring the ECG data. From the corrected ECG signal, we derived IBI, HR, and RMSSD for each experiment phase. In all studies, RMSSD was ln transformed because of skewed data distributions.

2.4.1. Belief in experimental VR setup

During the debriefing procedure of the VR-TSST, experimenters rated the participants as ‘believer’ (participant believed the entire experimental set-up), ‘doubter’ (e.g., participant questioned whether she was talking to real people, whether the audience members were actual professors), or ‘non-believer’ (participant was quite certain that pre-recorded messages were used). Participants were additionally asked to indicate how certain they were that they were talking to ‘real’ people during the task (0 to 100%).

2.5. Statistical analysis

Descriptive statistics comprised means and standard deviations for continuous variables, and frequencies for categorical variables. Pearson Chi-square tests were used to compare the subset samples on categorical sample characteristics (i.e., sex, smoking), while univariate analyses of variance (ANOVA) was used to assess differences on continuous variables (i.e., age, BMI). Specific to the VR sample, repeated measures ANOVA was used to assess the effect of believing the VR setup on the emotional and physiological stress response. Specific to the two *in vivo* samples, frequencies were calculated for adherence to health behavior guidelines, and chi square tests gauged potential differences in these percentages adherence.

The STAI-S (VR) and the emotion questionnaire (*in vivo*, control) were summed for the resting baseline and the stress score. Because the scale of the self-reported anxiety measures differed per study, the anxiety scores were first standardized around their resting mean (SD). Then the standardized scores were merged into one comparable score.

Data analysis for the physiological measures was as follows: As a manipulation check, we first examined the general reactions to the virtual TSST by testing the within-person time effect in an otherwise unadjusted analysis. The RMSSD variables were log-transformed, because these variables had skewed distributions (Shapiro-Wilk <0.05).

To compare the effects of the three TSST types on anxiety change and cardiovascular reactivity, a series of MIXED linear models were conducted, with anxiety (2 time levels: rest - stress), inter-beat interval and RMSSD (4 time levels) as outcome measures respectively. Time (Baseline, Preparation, Speech, Math) was entered as the repeated measures variable with an unstructured covariance matrix. For all models we tested whether the models improved when adding a random intercept. TSST type was entered as a fixed factor, as was Time. For each Model, we first tested the main effects of TSST type and Time, and their interaction. A significant interaction would indicate that the TSST induced physiological and emotional reactivity profile differed by TSST type. Then, we tested the significance of a random intercept, and finally sex was included as a covariate in a second step because of its established effects on emotion and physiology. We tested whether sex was a significant addition to the model using the AIC relative likelihood calculations. As our hypothesis was about equivalence of the TSST versions, we performed a TOST equivalence test for independent samples, based on Welch’s *t*-test (Lakens, 2017), when TSST type rendered a non-significant effect in the MIXED linear modeling. For this TOST equivalence test, we need to set equivalence boundaries. We followed the guidance of (Lakens, 2017), and chose the smallest effect sizes we had statistical power for to detect as equivalence boundaries, i.e. Cohen’s *d* of 0.40/–0.40.

Two-sided *p*-values are reported and a two-sided *p*-value <.05 was used for hypothesis testing. All analyses were conducted in SPSS (version 24).

3. Results

3.1. Participants & manipulation check

A total of 178 participants participated in either the *in vivo* TSST ($N = 106$; 51 women), the VR-TSST ($N = 52$), or the control TSST ($N = 20$).

Participant characteristics are displayed in Table 1. There were significant sex differences between the three samples, while no significant differences were found in age and health behaviors. In the VR-TSST, 19 participants (39.6%) expressed some doubts about the experimental setup, whereas fifteen (31.3%) were rated as ‘believers’ and fourteen (29.2%) as ‘non-believers’ (1 missing). Participants were on average 54.9% sure that they were talking to ‘real’ people, with $SD = 26.9$ and answers ranged from 0 to 100%. Believing the VR setup was unrelated to the anxiety response to the VR-TSST ($p = .484$, partial $\eta^2 = 0.03$), but was related to the physiological response ($F_{IBI}(6, 129) = 3.05, p = .008$, partial $\eta^2 = 0.12$; $F_{RMSSD}(6, 129) = 1.41, p = .218$, partial $\eta^2 = 0.06$), with believers showing a larger heart rate response to stress than doubters/non-believers.

There were no significant differences between the virtual TSST versus the *in vivo* TSST and control TSST (PHEMORE study) participants in their adherence to the health behavior guidelines prior to study participation (Ps between 0.32 and 0.57).

Examining the three data subsets separately, we first established whether the three versions of the TSST elicited the expected emotional and physiological responses. With respect to anxious mood, we found a significant response to the VR-TSST ($F(1, 50) = 161.59, p < .001$; partial $\eta^2 = 0.76$) and the *in vivo* TSST ($F(1, 105) = 238.17, p < .001$; partial $\eta^2 = 0.69$), while there was no response in the control TSST ($F(1,19) = 1.39, p = .25$; partial $\eta^2 = 0.07$). Results also showed that all versions of the TSST elicited significant changes in heart period ($F_{VR-TSST}(3, 144) = 85.77, p < .001$, partial $\eta^2 = 0.64$; $F_{in vivo}(3, 312) = 131.73, p < .001$, partial $\eta^2 = 0.56$, $F_{control}(3, 51) = 14.86, p < .001$, partial $\eta^2 = 0.47$). Repeated contrast analysis showed that all subsequent time points significantly differed from each other ($p < .001$). The VR-TSST ($F(3, 144) = 18.03, p < .001$; partial $\eta^2 = 0.27$) and the *in vivo* TSST ($F(2.4, 247.9) = 30.28, p < .001$; partial $\eta^2 = 0.23$) also elicited significant reductions in RMSSD, while the control TSST did not ($F(3, 51) = 1.69, p = .18$; partial $\eta^2 = 0.09$). Contrast analysis showed that while for the VR-TSST preparation and speech were not significantly different from each other ($p = .78$), the other time points were ($p < .001$). For the *in vivo* TSST all subsequent time points differed significantly ($p < .05$).

3.2. Comparing the TSST versions

3.2.1. Anxious mood

The mixed linear model with two levels of anxious mood as an outcome measure, a random intercept (see online results supplement for modeling results), two main effects (Time, TSST type) and their interaction, showed that the emotional response to the VR-TSST was not different from the emotional response to the *in vivo* TSST ($F(1,119) = 0.216, p = .643$; H1). The response size was 2.17 (se = 0.17) standardized units in the VR-TSST, 2.05 (se = 0.17) standardized units in the *in vivo* TSST and 0.034 (se = 0.24) in the control TSST (Fig. 2). The equivalence test (TOST) indicated that the observed effect size for

Table 1
Sample characteristics.

	VR-TSST (N = 52)	<i>In vivo</i> TSST (N = 106)	Control TSST (N = 20)	<i>p</i> value
	N(%) or mean ± s.d.	N(%) or mean ± s.d.	N(%) or mean ± s.d.	
Sex (women)	52 (100%)	68 (64%)	8 (40%) ^a	<.001
Age (years)	19.9 ± 1.8	20.5 ± 2.8	21.2 ± 1.9	.701
Smoking (yes)	12% (6)	15% (16)	20% (4)	.644
BMI (kg/m ²)	22.7 ± 4.5	22.0 ± 3.3	22.5 ± 3.0	.496

Note: results are presented as % (n), unless otherwise indicated. Column proportions were compared with the Fisher exact test. A subscript letter (a, b, c) attached to the percentages indicates whether samples are all different from each other, or that one sample stood out (a, b).

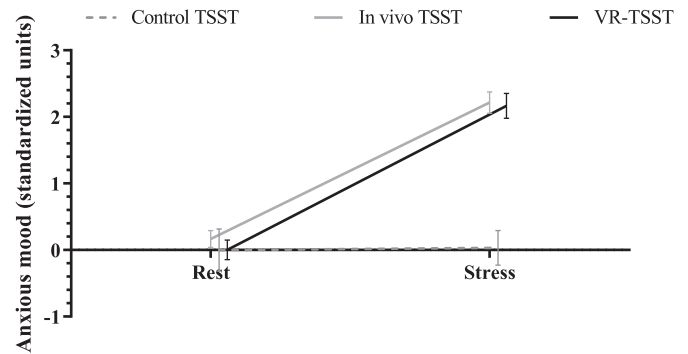


Fig. 2. Emotional response to the TSST stratified by TSST type. Note: This figure shows the standardized mean (se) level of anxious mood for the three TSST types during rest and during stress.

anxiety reactivity ($d = 0.09$) was significantly within the equivalent bounds of $d = -0.40$ and $d = 0.40$ ($t(114.95) = -1.71, p = .045$), which leads to the conclusion that for anxiety reactivity the VR-TSST is equivalent to the *in vivo* TSST.

Using the sample for hypothesis 2 (*i.e.* men and women, without the VR-TSST participants), the control TSST did not show an anxiety response (Time effect: $F(1,20) = 1.45, p = .243$). The *in vivo* TSST induced a significantly more pronounced anxiety response than the control TSST (Time by TSST version: $F(1, 126) = 36.65; p < .001$; H2). There was a significant main effect of sex, with women scoring on average 0.41 standardized units higher at rest and stress than their male counterparts ($F(1,126) = 5.00, p = .027$). Adding sex to the model did not affect the effect of TSST version (while sex was relevant to the model).

3.2.2. Physiology

Fig. 3 displays the physiological responses to the three versions of the TSST. None of the models included a random intercept, as these models provided a worse fit to the data (Online Results supplement). To test hypothesis 1 (VR-TSST not being different from *in vivo* TSST), mixed linear modeling with IBI (*i.e.* heart period) as an outcome measure showed a significant interaction between time and TSST type, indicating that the heart period response to the TSST differed per TSST version ($F(3; 117) = 9.22, p < .001$). Residuals were normally distributed. The *in vivo* TSST induced a larger heart period reduction than the VR-TSST (Fig. 3). Custom hypothesis testing for the difference in heart period response between the VR-TSST with the *in vivo* TSST (H1) revealed that the heart period response deviated during rest (*i.e.*, the VR-TSST participants were more relaxed; $\Delta IBI = -52.76; t = -2.68, p = .008$), and during the math stressor ($\Delta IBI = -73.80; t = -3.47, p = .001$), with the heart period being shorter (*i.e.* higher heart rate) in the *in vivo* TSST (Fig. 3). The VR-TSST IBI response was equivalent for the preparation ($t = -0.43, p = .669$) and speech periods ($t = -1.23, p = .222$).

For the second hypothesis (the *in vivo* TSST shows larger responses than the control TSST), mixed linear modeling showed that the IBI response differed in level between the *in vivo* and control TSST ($F(1, 124) = 4.81, p = .03$), and that the profile over time differed in some respects ($F(3, 124) = 2.34, p = .077$). Custom hypothesis testing of the interaction effect showed that in particular the profile of the IBI response differed from the control profile in two aspects: the change from rest to preparation ($t = 1.70, p = .091$) and the response to the math task ($t = 2.64, p = .009$). Adding sex as a covariate, though a relevant addition to the model (Online results supplement), did not result in a significant alteration of the results.

3.2.3. RMSSD

The mixed linear modeling with RMSSD as outcome measure showed an interaction between time and TSST type ($F(3, 133.98) = 3.32, p =$

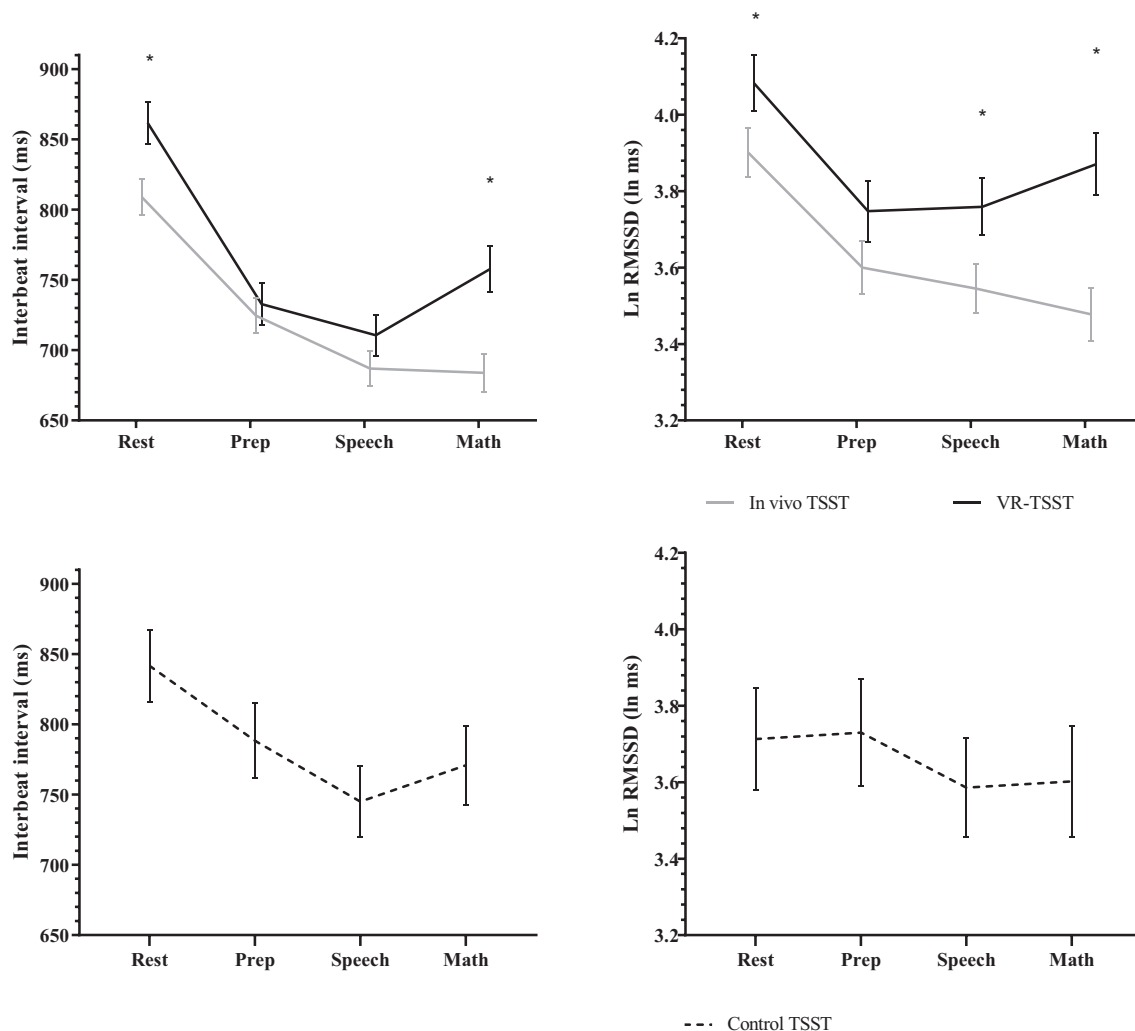


Fig. 3. Physiological TSST response profiles stratified by TSST type.

Note: Black lines represent the VR-TSST response profile. Solid gray line represent the response profile to the *in vivo* TSST, while dashed gray lines indicate the control TSST response profile. Prep = preparation period, ln = natural log; ms = milliseconds, TSST = Trier Social Stress Test. Error bars represent ± 1 SEM. * indicates significant contrast in interaction between time and TSST type (H1).

.022), indicating that the version of TSST significantly affected the RMSSD response profile. Custom hypothesis testing for the effects of this interaction showed that this was in particular the case for the speech response ($t = -2.17$, $p = .032$) and the math response ($t = -3.69$, $p < .001$), with the *in vivo* TSST inducing more parasympathetic withdrawal (Fig. 3).

With respect to the second hypothesis, Fig. 3 (bottom right figure in comparison to upper right figure) shows the RMSSD profile for the control group is lying in between the VR-TSST and the *in vivo* TSST response. What is remarkable, is that the control TSST participants overall show less parasympathetic activation, also in rest. Mixed linear modeling showed that the RMSSD response in the control TSST was equivalent to the *in vivo* TSST ($F(1, 120.49) = 1.93$, $p = .129$). Adding sex to the model as a covariate, though a relevant contribution to the model (online results supplement), did not change the effect of Time by TSST type. Sex was a non-significant covariate ($p = .23$).

4. Discussion

The present study examined whether a VR version of the TSST poses a viable alternative to the contemporary face-to-face TSST performed *in vivo* regarding the emotional and autonomic cardiac response. Overall, the results indicate that the emotional responses to the VR and *in vivo*

TSST were equivalent and both versions elicited higher responses than the control TSST. There were significant differences between TSST versions regarding the autonomic cardiac response, with less parasympathetic withdrawal and a smaller heart period stress response in the VR-TSST compared to the *in vivo* TSST. With respect to the second hypothesis, the heart period stress response was significantly larger than in the control TSST, while the parasympathetic withdrawal was equivalent. Together, these findings suggest that the VR-TSST elicits similar levels of negative affect, but less autonomic nervous system activation than the standard *in vivo* TSST.

Previous studies employing a VR-TSST have shown significant emotional (Fallon et al., 2016), neuroendocrine (Fallon et al., 2016; Jonsson et al., 2010; Ruiz et al., 2010) and cardiovascular (Jonsson et al., 2010; Kotlyar et al., 2008) responses, suggestive of successful production of the acute stress response. However, these studies did not make a direct comparison between responses to a VR type stress test with *in vivo* tests. A few previous studies have made a direct comparison regarding emotional, neuroendocrine, and cardiovascular responses (Kelly et al., 2007; Kothgassner et al., 2016; Shiban et al., 2016; Zimmer et al., 2019), and all of these studies used a head mounted display for VR presentation. With respect to the emotional stress response, our findings showed equal efficiency in producing an emotional (*i.e.* anxiety) stress response to the VR- and *in vivo* TSST. This is consistent with previous

work showing no differences in perception of stressfulness, appraisal of stress, or responses of anxiety between VR and *in vivo* TSSTs (Kelly et al., 2007; Kothgassner et al., 2016; Shibana et al., 2016; Zimmer et al., 2019).

Differences do exist between studies with respect to the comparability of the autonomic cardiac response. Our VR-TSST produced a smaller heart period response, particularly during math, and less parasympathetic withdrawal compared to the *in vivo* TSST (the VR math task may not have been challenging enough, which is discussed later in this article). One prior study that directly compared VR-TSST with *in vivo* TSST also showed an equivalent parasympathetic withdrawal (Kothgassner et al., 2016). Our findings concur with the recent study of Zimmer and colleagues reporting lower heart rate responses in the VR-TSST condition, as compared to *in vivo* (Zimmer et al., 2019). However, our findings are not in line with other immersive VR vs. *in vivo* comparison studies that have shown no differences in cardiovascular reactivity between conditions (Kothgassner et al., 2016; Shibana et al., 2016). In addition, given the potential role of the level of experienced immersion, it is of note that Kothgassner and colleagues used a group audience in their study, which may have influenced their results as well (Kothgassner et al., 2016). Increases in heart rate may be considered as an indirect measure of task engagement (Seery, 2011). The reduced capacity to mount a heart rate/period response observed in the current study may be associated with the level of task engagement, but also with the believability of the task. Zimmer et al., who also found a slightly attenuated heart rate response to the VR-TSST (Zimmer et al., 2019), suggested the TSST may be difficult to replicate in a virtual environment due to its conceptualization as a socially evaluated and uncontrollable performance stressor. It is of note that the anticipation response, which is a private, passive response, prior to the active performance tasks was similar in the VR-TSST and the *in vivo* TSST. The social evaluation and negative feedback during the active performance stressors may be less believable in VR. Our own findings showed that individuals who believed the experimental set-up showed increased heart rate responses in the VR setting compared to others who believed the VR setting to a lesser extent or not at all. This suggests that improving the believability of the VR-TSST may also affect its effectiveness. Engagement with the tasks at hand is also an important determinant of the responses to the *in vivo* TSST (Seery, 2011), which makes believability also an important aim in *in vivo* tasks. Future studies may want to examine whether making participants believe in the experimental set-up may be one way to further increase the effectiveness of VR-TSSTs, possibly even to the level of *in vivo* TSST. Improvements to the VR-TSST could be achieved by recording more voice messages that can be used to better, and more flexibly, simulate conversations between the participant and avatars.

Believability of the VR-TSST may also be improved by increasing the level of immersion into the virtual task by making the visual images more realistic than the current avatars. A recent meta-analysis showed that immersive VR-TSSTs are more effective in inducing a cortisol response, compared to non-immersive TSSTs, such as the currently used Second-life screen version (Helminen et al., 2019). However, in immersive TSSTs participants may realize that they are not actually presenting in front of a real audience, but for programmed avatars, which may set limits to effective stress induction. Moreover, head mounted displays and CAVE environments may cause nausea and simulation sickness in some participants (Pan and Hamilton, 2018). These observations indicate that the level of immersion, believability of the VR setting, and task engagement all may be important moderators of the physiological response to a virtual stress task. It is important to quantify their role in VR-TSST reactivity in future research.

It is also unclear if the physical presence of an (evaluative) audience contributes to physiological reactivity in a VR-TSST, regardless of the social evaluative threat and negative feedback. Dickerson et al. (2008) found that negative social evaluation, but not mere social presence, elicits a neuroendocrine response to a laboratory stressor. Our data are in accordance in this respect, as in the control TSST the social evaluative aspect was absent (though there was social presence) and emotional and

heart rate responses to the control TSST were attenuated. Additionally, it has been shown in previous research that in the *in vivo* TSST, social evaluation and audience size do matter. Anxiety, cortisol and autonomic activation all have shown increased reactivity when the task had a socially evaluative character. Moreover, physiological reactivity increased in parallel with increasing audience size (Bosch et al., 2009). While our VR-TSST had a two-person audience, physiological responses were smaller than those of the *in vivo* TSST. Using a larger audience thus may also increase response sizes.

Considering our findings, the math task of the current VR-TSST could be improved. The math task was substantially less stressful (*i.e.* less physiological arousal) than the *in vivo* TSST. Most likely, this may be attributed to differences in the negative feedback from evaluators (*i.e.* number and timing flexibility of interruptions, facial expression, and tone of voice) and the associated social evaluative threat experienced by the participants. Social evaluative threat is most likely to occur when failure or poor performance could reveal lack of a valued trait or ability. It is a key contributor to the physiological stress response in the TSST (Dickerson and Kemeny, 2004). The current VR math task can thus be improved by intensifying and more flexibly applying the negative feedback (gestures and comments) from evaluators. Examining the scripted text of evaluators in the *in vivo* TSST and the VR-TSST, it is evident that the instruction of the math task in the *in vivo* TSST already contained more evaluative primers. Furthermore, the comments during the performance were politer and nicer (*i.e.*, “little slower”, “please try to”, ...) than the *in vivo* TSST. These are clear improvements that need to be made to the VR math task.

The results of the current study should be viewed in light of several limitations and strengths. We did not randomize participants into any of the three TSST arms, but rather used separate samples of studies that were executed in the same lab, with a similar TSST overall design, though there were slight protocol variations. Because of this merge, we also needed to standardize the scale of the anxiety responses of the *in vivo*/control TSST study around their baseline mean. It should be noted that differences in the tools to assess anxiety may have introduced bias. The equipment to record the physiological measures also differed between the two studies, but it is unlikely to have affected the heart rate and RMSSD findings. In addition, while the *in vivo* TSST had a control TSST counterpart, there was no control condition for the VR-TSST, which would be a suggestion for future research. Because of the overall study design, the VR-TSST only was performed in women, while the other two TSST protocols were performed in both women and men. Since our VR-TSST only included women, we cannot conclusively say the TSST responses were equivalent for men and women, regardless of TSST type. Future studies examining sex differences will be important. We did not have any performance measures (*e.g.* score on the math test), to compare between the VR-TSST and the *in vivo* TSST, which may provide some more detail. Nevertheless, the *in vivo* TSST elicited a stronger physiological response. The difference in physiological response may be explained by a more lenient math test protocol in the VR-TSST, and VR-related issues discussed above. Another limitation is that there was no *a priori* power analysis (convenience comparison), and statistical analyses were not adjusted for multiple testing. However, given the sample size, strict corrections of the alpha level were not possible. Finally, because the VR-TSST participants were in the placebo group of a larger trial, the placebo administration could have led to attenuated responses when participants thought they were given oxytocin. We tested this in a post-hoc analysis, and no differences were found, which adds confidence to our findings. Strengths of the study included the relatively large sample size, the combined assessment of emotional and physiological reactivity, and the inclusion of sex as a covariate.

In conclusion, the present findings demonstrate that social evaluative stress induced in a screen-based VR setting produced similar emotional, and somewhat attenuated autonomic cardiovascular responses as compared to *in vivo*. We recommend intensifying the social-

evaluative threat and time pressure during the math task by altering and maximizing interaction in SecondLife, and increasing audience size, which would be expected to lead to larger physiological responses. Moreover, our findings indicate that belief in the experimental-set up results in a more effective stress induction. Thus, the credibility of the experimental set-up of VR-TSSTs may be one important, but often neglected, moderating factor that could increase the effectiveness of VR-TSST.

Data availability

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

Study 1 was sponsored by the Tilburg University Alumni Fund, the Department of Medical and Clinical Psychology, and the Center of Research on Psychological and Somatic disorders, Tilburg University, The Netherlands. The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijpsycho.2021.01.010>.

References

- Allen, A.P., Kennedy, P.J., Cryan, J.F., Dinan, T.G., Clarke, G., 2014. Biological and psychological markers of stress in humans: focus on the Trier Social Stress Test. *Neurosci. Biobehav. Rev.* 38, 94–124. <https://doi.org/10.1016/j.neubiorev.2013.11.005>.
- Bosch, J.A., de Geus, E.J., Carroll, D., Goedhart, A.D., Anane, L.A., van Zanten, J.J., Edwards, K.M., 2009. A general enhancement of autonomic and cortisol responses during social evaluative threat. *Psychosom Med* 71 (8), 877–885. <https://doi.org/10.1097/PSY.0b013e3181baef05>.
- Cohen, S., Janicki-Deverts, D., Miller, G.E., 2007. Psychological stress and disease. *JAMA* 298 (14), 1685–1687. <https://doi.org/10.1001/jama.298.14.1685>.
- De Geus, E.J., Willemsen, G.H., Klaver, C.H., Van Doornen, L.J., 1995. Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. *Biol. Psychol.* 41 (3), 205–227.
- Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychol. Bull.* 130 (3), 355–391. <https://doi.org/10.1037/0033-2909.130.3.355>.
- Dickerson, S.S., Mycek, P.J., Zaldivar, F., 2008. Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. *Health Psychol.* 27 (1), 116–121. <https://doi.org/10.1037/0278-6133.27.1.116>.
- Duijndam, S., Karreman, A., Denollet, J., Kupper, N., 2020. Physiological and emotional responses to evaluative stress in socially inhibited young adults. *Biol Psychol.* 107811. <https://doi.org/10.1016/j.biopsycho.2019.107811>.
- Fallon, M.A., Careaga, J.S., Sbarra, D.A., O'Connor, M.F., 2016. Utility of a virtual Trier Social Stress Test: initial findings and benchmarking comparisons. *Psychosom. Med.* 78 (7), 835–840. <https://doi.org/10.1097/PSY.0000000000000338>.
- 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>.
- Grinberg, A. M., Careaga, J. S., Mehl, M. R., & O'Connor, M.-F. (2014). Social engagement and user immersion in a socially based virtual world. *Computers in Human Behavior*, 36, 479–486. doi:10.1016/j.chb.2014.04.008.
- Guadagno, R., Blascovich, J., Bailenson, J., McCall, C., 2007. Virtual humans and persuasion: the effects of agency and behavioral realism. *Media Psychol.* 10, 1–22. <https://doi.org/10.1080/15213260701300865>.
- 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>.
- 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>.
- Jonsson, P., Wallergard, M., Osterberg, K., Hansen, A.M., Johansson, G., Karlson, B., 2010. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: a pilot study. *Psychoneuroendocrinology* 35 (9), 1397–1403. <https://doi.org/10.1016/j.psyneuen.2010.04.003>.
- Kelly, O., Matheson, K., Martinez, A., Merali, Z., Anisman, H., 2007. Psychosocial stress evoked by a virtual audience: relation to neuroendocrine activity. *CyberPsychol. Behav.* 10 (5), 655–662. <https://doi.org/10.1089/cpb.2007.9973>.
- 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>.
- Knowles, L.M., Stelzer, E.-M., Jovel, K.S., O'Connor, M.-F., 2017. A pilot study of virtual support for grief: feasibility, acceptability, and preliminary outcomes. *Comput. Hum. Behav.* 73, 650–658. <https://doi.org/10.1016/j.chb.2017.04.005>.
- Kothgassner, O.D., Felnhöfer, A., Hlavacs, H., Beutl, L., Palme, R., Kryspin-Exner, I., Glenk, L.M., 2016. Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. *Comput. Hum. Behav.* 62, 124–135. <https://doi.org/10.1016/j.chb.2016.03.081>.
- Kotlyar, M., Donahue, C., Thuras, P., Kushner, M.G., O'Gorman, N., Smith, E.A., Adson, D.E., 2008. Physiological response to a speech stressor presented in a virtual reality environment. *Psychophysiology* 45 (6), 1034–1037. <https://doi.org/10.1111/j.1469-8986.2008.00690.x>.
- Kupper, N., Pelle, A., Denollet, J., 2013. Association of Type D personality with the autonomic and hemodynamic response to the cold pressor test. *Psychophysiology* 50 (12), 1194–1201. <https://doi.org/10.1111/psyp.12133>.
- Kupper, N., Jankovic, M., Kop, W.J., 2020. Individual differences in cross-system physiological activity at rest and in response to acute social stress; the PHEMORE study. *Psychosom. Med.* <https://doi.org/10.1097/PSY.0000000000000901> (in press).
- Lakens, D., 2017. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* 8 (4), 355–362. <https://doi.org/10.1177/1948550617697177>.
- Marteau, T.M., Bekker, H., 1992. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *Br. J. Clin. Psychol.* 31 (3), 301–306.
- Montero-Lopez, E., Santos-Ruiz, A., Garcia-Rios, M.C., Rodriguez-Blazquez, R., Perez-Garcia, M., Peralta-Ramirez, M.I., 2016. A virtual reality approach to the Trier Social Stress Test: contrasting two distinct protocols. *Behav. Res. Methods* 48 (1), 223–232. <https://doi.org/10.3758/s13428-015-0565-4>.
- Olufsen, M. S., Ottesen, J. T., Tran, H. T., Ellwein, L. M., Lipsitz, L. A., & Novak, V. (2005). Blood pressure and blood flow variation during postural change from sitting to standing: model development and validation. *Journal of applied physiology* (Bethesda, Md.: 1985), 99(4), 1523-1537. doi:<https://doi.org/10.1152/japophysiol.00177.2005>.
- Pan, X., Hamilton, A.F.C., 2018. Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. *Br. J. Psychol.* 109 (3), 395–417. <https://doi.org/10.1111/bjop.12290>.
- Riem, M.M., Kunst, L.E., Bekker, M.H.J., Fallon, M.A., Kupper, N., 2020. Intranasal oxytocin enhances stress-protective effects of social support in women with childhood adversity during a Virtual Trier Social Stress Test. *Psychoneuroendocrinology* 111.
- Ruiz, A.S., Peralta-Ramirez, M.I., Garcia-Rios, M.C., Muñoz, M.A., Navarrete-Navarrete, N., Blazquez-Ortiz, A., 2010. Adaptation of the Trier Social Stress Test to virtual reality: psycho-physiological and neuroendocrine modulation. *J Cyber Ther Rehabil.* 3 (4), 405–415.
- Schultze, U., 2010. Embodiment and presence in virtual worlds: a review. *JIT* 25, 434–449. <https://doi.org/10.1057/jit.2010.25>.
- Seery, M.D., 2011. Challenge or threat? Cardiovascular indexes of resilience and vulnerability to potential stress in humans. *Neurosci. Biobehav. Rev.* 35 (7), 1603–1610. <https://doi.org/10.1016/j.neubiorev.2011.03.003>.
- Shiban, Y., Diemer, J., Brandl, S., Zack, R., Muhlberger, A., Wust, S., 2016. Trier Social Stress Test in vivo and in virtual reality: dissociation of response domains. *Int. J. Psychophysiol.* 110, 47–55. <https://doi.org/10.1016/j.ijpsycho.2016.10.008>.
- Wilson, C., Soranzo, A., 2015. The use of virtual reality in psychology: a case study in visual perception. *Computational and Mathematical Methods in Medicine* 2015, 151702. <https://doi.org/10.1155/2015/151702>.
- Zimmer, P., Buttler, 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>.