Psychoneuroendocrinology 132 (2021) 105342

Contents lists available at ScienceDirect

Psychoneuroendocrinology

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

Stability and test-retest reliability of different hormonal stress markers upon exposure to psychosocial stress at a 4-month interval

Ann-Kathrin Kexel [a,*], Bruno Kluwe-Schiavon [a], Monika Visentini [a], Leila M. Soravia [b], Clemens Kirschbaum [c], Boris B. Quednow [a,d,*]

[a] Experimental and Clinical Pharmacopsychology, Department of Psychiatry, Psychotherapy, and Psychosomatics, Psychiatric Hospital of the University of Zurich, Lenggstrasse 31, 8032 Zurich, Switzerland
[b] Translational Research Center, University Hospital of Psychiatry, University of Bern, Bolligenstrasse 111, 3000 Bern 60, Switzerland
[c] Department of Biopsychology, Technical University Dresden, Zellescher Weg 19, 01069 Dresden, Germany
[d] Neuroscience Center Zurich, University of Zurich and Swiss Federal Institute of Technology Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

ARTICLE INFO

Keywords:
Habituation
Hypothalamus-pituitary-adrenal axis reactivity
Cortisol
Subjective stress
Trier social stress test

ABSTRACT

The Trier Social Stress Test (TSST) has been shown to reliably induce physiological stress responses in the hypothalamus-pituitary-adrenal (HPA) and in the sympathetic-adrenal-medullary (SAM) axis in cross-sectional studies. However, it was also reported that repeated exposure to the TSST might be associated with habituation, mainly of the HPA axis responsivity. Thus, in all longitudinal stress studies involving repeated TSST administration, potential habituation of the HPA axis response complicates the interpretation of results. The goal of the present study was therefore to assess stability and test-retest reliability of a number of different endocrinological stress markers as well as subjective stress responses after two exposures to the TSST four months apart. We assessed salivary and plasma cortisol profiles, plasma ACTH and noradrenaline profiles, as well as subjective stress ratings in healthy volunteers before, during, and after the TSST at six time-points both at test-day 1 (TSST_1, $n = 42$) and test-day 2 (TSST_2, $n = 34$) 4-months later. Half of the participants received the TSST in the early, the other half in the late afternoon. Discontinuous growth models were applied to model three phases of the stress response (preTSST, reactivity, recovery) for each marker. Subsequently, the stability of these phases was analyzed. Stability and test-retest reliability of standard physiological stress markers such as Area-under-the-Curve (AUC_G, AUC_I), Absolute Peak Change, and Relative Peak Change (RPC) were analyzed as well. We did not observe strong test-retest effects in any of the endocrinological measures. In contrast, test-retest effects in subjective stress were characterized by a faster drop directly after the second TSST, whereas the initial increase before the test period was the same for both test-days. Regarding test-retest-reliability, AUC_G was the most reliable measure across all endocrinological and subjective stress markers (range: $r = .606$ to .858), while AUC_I and RPC (range: $r = -.146$ to .548) were least reliable. A 4-month interval is a sufficient time interval between two repeated TSST exposures to largely reinstate the physiological stress response, which was also true for the initial psychological stress response. Thus, the TSST is well applicable in longitudinal studies.

1. Introduction

The Trier Social Stress Test (TSST; Kirschbaum et al., 1993) is one of the most widely used tools in the field of psychological stress research to experimentally induce acute psychosocial stress. It is characterized as a motivated performance task with high levels of social-evaluative threat and uncontrollability, two elements that are particularly suited to elicit hypothalamus-pituitary-adrenal (HPA) axis responses (Dickerson and Kemeny, 2004). Thus, the TSST has been shown to reliably influence key elements of the stress response system, inter alia, the HPA axis and the sympathetic-adrenal-medullary (SAM) axis (for a review see Allen et al., 2014). For instance, two- to three-fold increases in salivary cortisol levels have been observed in 70–80% of participants (Kirschbaum et al., 1993; Kudielka et al., 2007) and one- to three-fold increases have been reported in adrenocorticotropic hormone (ACTH) levels (Foley and Kirschbaum, 2010). Regarding the SAM axis, elevations in adrenaline, noradrenaline, and salivary alpha-amylase in response to the TSST have been found (e.g., Gold et al., 2004; Jezova et al., 2004; Rohleder et al.,

* Correspondence to: University Hospital of Psychiatry, Experimental and Clinical Pharmacopsychology, Lenggstrasse 31, CH-8032 Zurich, Switzerland.
E-mail addresses: ann-kathrin.kexel@bli.uzh.ch (A.-K. Kexel), quednow@bli.uzh.ch (B.B. Quednow).

https://doi.org/10.1016/j.psyneuen.2021.105342
Received 10 February 2021; Received in revised form 21 May 2021; Accepted 21 June 2021
Available online 25 June 2021

2004; Thoma et al., 2012). In addition, individuals do not only show a physiological stress response, but also report increased levels of perceived subjective stress (e.g., Hellhammer and Schubert, 2012; Sugaya et al., 2012).

It was demonstrated that repeated exposure to the TSST can lead to habituation of the HPA axis response, i.e., a decrease in the response magnitude of salivary and plasma cortisol as well as ACTH (e.g., Federenko et al., 2004; Gerra et al., 2001; Kirschbaum et al., 1995; Schommer et al., 2003; Wüst et al., 2005). This has been shown for varying test intervals, from one day (e.g., Höhne et al., 2014; Kirschbaum et al., 1995) to seven days (e.g., Federenko et al., 2004; Gerra et al., 2001; Wüst et al., 2005), and four weeks (e.g., Schommer et al., 2003; Wolf et al., 2002). While the majority of individuals display HPA axis habituation, a smaller group does not show any changes in their stress response over repeated exposure (Gerra et al., 2001; Kirschbaum et al., 1995; Schommer et al., 2003; Wüst et al., 2005) and a minority actually experiences sensitization (Kudielka et al., 2006; Wüst et al., 2005). According to Manigault et al. (2020), a high HPA axis reactivity during the first TSST is associated with habituation during repeated TSST exposure, whereas low reactivity during the first TSST is actually associated with subsequent sensitization. Furthermore, greater subsequent HPA axis habituation is linked, inter alia, to trait cognitive reappraisal (Roos et al., 2019), fewer rumination after the initial TSST (Gianferante et al., 2014), and a basal cortisol activity that is indicative of better health outcomes (Chen et al., 2017). However, at group-level, habituation generally occurs at short-term follow-ups. In contrast, test-retest effects after long-term follow-ups have only scarcely been investigated and are therefore still not well understood. Moreover, habituation to the TSST seems to be rather specific to the HPA axis as no response changes could be observed in the SAM system (Gerra et al., 2001; Schommer et al., 2003). However, habituation in state anxiety (Federenko et al., 2004; Jönsson et al., 2015, 2010) and reduced self-reported negative affect (McInnis et al., 2015) in response to repeated TSST exposure have been observed.

Habituation of the stress response is generally considered as an adaptive mechanism to reduce allostatic load (McEwen, 1998; McEwen and Stellar, 1993). Repeated performance of the TSST is therefore a useful tool to investigate altered habituation patterns as a vulnerability factor for stress-related diseases and associated factors that might contribute to disrupted habituation. However, if researchers are interested in testing stress responses after interventions or in longitudinal relationship to naturally occurring changes over time, HPA axis habituation complicates interpretation of results as effects of, e.g., treatment, can hardly be separated from habituation in pre- and post-treatment measures. It has been proposed that slight changes in the TSST protocol, such as changing the speech topic and arithmetic task to induce novelty (Kirschbaum et al., 1995), and longer time periods between tests (Foley and Kirschbaum, 2010) are sufficient to reduce habituation. However, only few published studies have addressed the latter. For instance, Petrowski et al. (2012) showed that salivary cortisol responses did not habituate after a 10-week interval. Boesch et al. (2014) did not test for HPA axis habituation but demonstrated self-reported affective habituation in the group version of the TSST after a 10-week interval. Yet, they only assessed pre- and post-TSST affect in men. As expected, changes in the SAM axis were not observed (Boesch et al., 2014). Furthermore, Asbrand et al. (2019) and Het et al. (2020) already used repeated TSST exposure in two intervention studies. Asbrand et al. (2019) used a 12-week interval to test the influence of cognitive-behavioral therapy on stress responses in salivary cortisol and alpha-amylase in children with social anxiety disorder. Het et al. (2020) used an average interval of eight weeks to assess the stress response as measured by salivary cortisol and alpha-amylase, heart rate, and heart rate variability in eating disorder patients before and after completion of an in-patient treatment program. Nonetheless, to the best of our knowledge, no study has yet investigated other HPA axis biomarkers other than salivary cortisol to test test-retest stability after long-term follow-ups. Moreover, no study has yet compared different endocrinological biomarkers of stress and the subjective stress response with each other after prolonged time intervals between two repeated TSST exposures. Thus, comprehensive reports on test-retest stability after long-term follow-ups are currently lacking.

We therefore exposed male and female individuals two times to the TSST and assessed salivary and plasma cortisol, ACTH, noradrenaline, and subjective stress repeatedly during both test sessions. We chose a 4-month follow-up interval as it was suggested to prevent cortisol habituation (Foley and Kirschbaum, 2010). Contrary to other test-retest studies, we kept the TSST protocol equal between test-days, as longitudinal studies usually require constant settings across all measurement points. We used discontinuous growth models, a variation of linear mixed models (LMMs), to analyze TSST trajectories. LMMs have only rarely been used to investigate habituation but are particularly suited due to the hierarchical data structure. Moreover, we not only looked at the trajectories but also calculated proposed TSST stress markers (AUC_G, AUC_I, Absolute, and Relative Peak Change) to assess test-retest stability.

Based on the current literature, we hypothesized that the TSST response would be stable between first and second exposure. Moreover, test-retest reliability for TSST stress markers was assessed in an exploratory fashion. Overall, this study is, until now, the most comprehensive and detailed study investigating TSST response stability and test-retest reliability for time intervals of more than four weeks.

2. Methods

2.1. Participants

54 individuals (22 females/ 32 males) participated in a screening session to verify in- and exclusion criteria. Inclusion criteria were being able to read, understand and provide written informed consent; German fluency; age between 18 and 52 years. Exclusion criteria were a severe neurological disorder or brain injury; a current diagnosis of an infectious disease or severe somatic disorder; a history of autoimmune, endocrine, and rheumatoid arthritis; intake of medication with potential action on the central nervous system during the last three days; a DSM-IV-R Axis I adult psychiatric disorder, or recurrent illegal substance use (> 15 occasions lifetime per substance, with the exception of cannabis use); a family history of genetically mediated psychiatric disorders ($h^2 > 0.5$, e.g., autism spectrum disorder, bipolar disorder, and schizophrenia); for women pregnancy, breastfeeding, or menstruation. Exclusion criteria were assessed at screening session using the Structured Clinical Interview-I for DMS-IV Axis I disorders (Wittchen et al., 1997) and the standardized Interview for Psychotropic Drug Consumption (Quednow et al., 2004) to determine self-reported substance use. Medical conditions putatively affecting physiological stress responses were assessed in a separate structured interview. Additionally, substance use during the last 4 months was objectively quantified by hair toxicology of a proximal 4 cm-hair-segment using liquid chromatography tandem mass spectrometry (LC-MS/MS). Moreover, we aimed to test women during the luteal phase of their menstrual cycle, but due to organizational constraints this was not always possible. However, women did not participate during menstruation.

After application of these criteria and counting dropouts, 42 individuals (16 females) remained for test-day 1 and 34 (12 females) for test-day 2 (for details see Supplements). Sample size considerations were based on previous research. Medium effect sizes for cortisol habituation were found with small samples ($N = 25$; Kudielka et al., 2006), and cortisol and ACTH responses were also observed in small samples in the TSST ($N = 20$; Kirschbaum et al., 1993) and in a study investigating social-evaluative threat ($N = 28$; Dickerson et al., 2008).

The study was approved by the Ethics Committee of the Canton Zurich (BASEC ID 2016-00278) and preregistered in the International Standard Randomized Controlled Trial Number Registry (ISRCTN10690316). All participants provided written informed consent

Table 1
Sample characteristics.

	TSST_1 sample (n = 42)	TSST_2 sample (n = 34)
Sex (m/f) (n)	26/16	22/12
Age	29.7 (6.9)	29.5 (6.3)
	(21 – 50)[a]	(21 – 45)
BMI	23.1 (3.2)	23.6 (3.2)
	(17.3 – 29.1)[a]	(17.3 – 29.1)
Verbal IQ	102.3 (9.1)	101.6 (9. 4)
Years of school education	10.4 (1.5)	10.5 (1.5)
Smoker/Non-Smoker (n)[b]	33/9	26/8
Cigarettes/week[c]	67.6 (43.5)	78.8 (50.4)
Alcohol grams/week	75.5 (71.9)	85.6 (82.6)
Early TSST/late TSST (n)	22/20	16/18
Menstrual cycle (n)[d]		
Follicular	5	5
Luteal	11	7
Hormonal contraception[d]	5	4

Note. Means and standard deviation of means in parenthesis.
[a] Range for age and BMI.
[b] Current smoker ($\geq$ 7 cigarettes/week).
[c] Only for smokers.
[d] Only for females.

in accordance with the Declaration of Helsinki. Sample characteristics are summarized in Table 1.

2.2. Procedure and study design

Individuals participated in the TSST at test-day 1 (TSST_1) and approximately four months (*median* = 4.3; *range* = 3.2–5.9) later at test-day 2 (TSST_2). As we aimed to investigate the effects of time-point, individuals underwent a counterbalanced TSST design varying the time of the psychosocial stress exposure. Thus, half of the individuals underwent the TSST in the early (between 01.30 pm and 02.45 pm) and the other half in the later afternoon (between 03.15 pm and 04.30 pm). The procedure was the same for both test-days (see also Supplements). The test sessions are schematically depicted in Fig. S1.

A detailed description of the TSST can be found in Kirschbaum et al. (1993), Kudielka et al. (2007), and Labuschagne et al. (2019). Briefly, the TSST consists of a preparation (10 min) and test period (10 min). In the beginning of the preparation period, individuals were introduced to the subsequent task and had then time to prepare a 5 min free speech on their suitability for a job of their choice. Subsequently, individuals were transferred to a second room where the TSST panel ('selection committee'; one male and one female confederate) was waiting for them. During the ensuing test period, individuals first gave their free speech (5 min) followed by a mental arithmetic task (5 min). Individuals were videotaped and voicerecorded during the test period. Confederates were unknown to the participants, remained neutral and restrained from any verbal or non-verbal feedback during the entire test period. Afterwards, individuals were escorted back to the first room. The TSST protocol was identical between TSST_1 and TSST_2, thus the conducted tasks remained the same. The TSST panel did not change for more than one person between TSST_1 and TSST_2. A debriefing of the TSST was done only at the end of TSST_2.

In the beginning of each test-day, participants drank 200 ml of juice concentrate with a high sugar content to standardize the blood glucose level before the measurement as it was suggested that the availability of glucose is a necessary requirement for the responsiveness of the HPA axis (Gonzalez-Bono et al., 2002; Kirschbaum et al., 1997; Kudielka et al., 2009). Blood and saliva samples ($T_{Beginning}$) were taken around 01:00 pm. For individuals in the early TSST condition, $T_{Beginning}$ samples were used as T_1 in the analysis of the psychosocial stress response (Fig. S1). Blood samples were drawn 25 (T_1; − 45 min) and 0 min (T_2; − 20 min) before the preparation period as well as 0 (T_3), 20 (T_4), 40 (T_5), and 65 min (T_6) after the test period. Saliva samples were taken 25 (T_1; − 45 min) and 0 min before (T_2; − 20 min) and after (T_3; − 10 min) the preparation period as well as 0 (T_4), 10 (T_5), 20 (T_6), and 40 min (T_7) after the test period. Blood samples were taken with BD Vacutainer® EDTA-tubes by a study nurse using an intravenous (i.v.) catheter placed in a forearm vein and immediately centrifuged. Plasma was aliquoted and stored at − 80 °C until analysis. Saliva samples were collected by Cortisol-Salivettes (blue cap, Sarstedt) and frozen directly after test-days at − 20 °C. We were not able to collect blood samples for one individual at both TSST_1 and TSST_2 and one individual at TSST_1 only, due to problems with placing the i.v. catheter.

2.3. Outcome measures

2.3.1. Endocrinological markers

ACTH, saliva and plasma cortisol were analyzed by immunoassays at Dresden LabService GmbH (Technical University of Dresden, Dresden, Germany). Noradrenaline was determined by high performance liquid chromatography. A detailed description of the used assays can be found in the test kits (ACTH: ACTH Elisa, IBL International GmbH, Hamburg, Germany; saliva cortisol: Cortisol Saliva Luminescence Immunoassay, IBL International GmbH, Hamburg, Germany; plasma cortisol: Cortisol Elisa, IBL International GmbH, Hamburg, Germany; noradrenaline: ClinRep® HPLC Complete Kit, Catecholamines in Plasma, Recipe Chemicals + Instruments GmbH, Munich, Germany). Interassay coefficients of variation for salivary cortisol were 5.8%, for plasma cortisol 7.2%, for ACTH 8.8%, and for noradrenaline 5.2%. Intraassay coefficients of variation for salivary cortisol were 4.3%, for plasma cortisol 3.5%, for ACTH 7.5%, and for noradrenaline 2.3%. Six individuals had single time-point missing data (1.3%) and three individuals had missing values at $\geq$ 3 time-points (2.2%) in noradrenaline. Handling of missing data is explained in the Supplements.

2.3.2. Subjective stress

Subjective stress was digitally assessed with an 11-point rating scale (*How stressed do you feel?*; 0 = not stressed, 10 = very stressed, with quarterly intervals) in the beginning of each test-day ($T_{Beginning}$), directly before the preparation (T_1) and test period (T_2), directly after the test period (T_3), and 65 min later (T_4).

2.4. Statistical analysis

2.4.1. Activation in the beginning of each test-day

LMMs (also known as linear multilevel models or random coefficient models) with a random-intercept for participant ID were used to assess baseline differences in cortisol, ACTH, noradrenaline, and subjective stress measures at the beginning of each test-day ($T_{Beginning}$). Test-day (dummy-coded as 0, TSST_1, and 1, TSST_2) was included as a fixed-effect.

2.4.2. Test-retest stability in TSST trajectories

Outliers were identified for each outcome measure separately and were defined as individuals with values larger than three times the interquartile range in the average value between the first and second sample that was taken in the beginning of each test-day, respectively (Jones, 2019). The first and second samples were chosen as individuals had not yet experienced the TSST. Three outliers were excluded from salivary, two for plasma cortisol, and one for ACTH (for details see Supplements). Salivary cortisol, ACTH, and noradrenaline were log-transformed prior to the analysis.

Discontinuous growth models (DGM; Singer and Willett, 2003) were used to analyze stability in TSST trajectories of salivary and plasma cortisol, ACTH, noradrenaline, and subjective stress. DGM are also referred to as piecewise hierarchical linear models (Hernández-Lloreda et al., 2004; Raudenbush and Bryk, 2002) and are a variation of LMMs. We fit 3-level DGM with individual samples (level-1) nested in test-days (TSST_1, TSST_2; level-2) and test-days nested in individuals (level-3).

Based on the combination of the known trajectories of the TSST stress response (e.g., Hellhammer and Schubert, 2012; Kirschbaum et al., 1999, 1993; Petrowski et al., 2010; Schommer et al., 2003) and the visual appearance of the descriptive trajectories at TSST_1, salivary cortisol and subjective stress response were divided into 3 linear components, and plasma cortisol, ACTH, and noradrenaline response into 4 linear components (details and coding schemes are described in the Supplements, Table S1). Time slopes for cortisol measures, ACTH, and noradrenaline were adapted for time and represent 10 min increments.

First, we fitted a null model. Second, we added interactions between test-day (dummy-coded as 0 for TSST_1, and 1 for TSST_2) and the respective linear time components as fixed-effects to investigate test-retest stability. Additionally, order (dummy-coded as 0 for early, and 1 for late TSST) was entered as a fixed-effect to test for possible effects of time-point. Third, to determine if order had an influence on the trajectories, we included interactions between order and the respective linear time components and compared this model with Bayesian's Information Criterion (BIC) to the simpler model fitted before. For all outcome measures, BIC was smaller for the simpler model without interactions between order and time components. Thus, interactions between order and linear time components were not included. Fourth, we tested random-slopes for time components by successively adding a random-slope for a time component to the simpler model. For plasma cortisol, ACTH, noradrenaline, and subjective stress models with a random-slope ran into convergence errors. Thus, random-slopes were not included for these outcome measures. BIC was used for all model comparisons.

To test the robustness of our results, sensitivity analyses were conducted by adding covariates to the models. Covariates considered were sex (coded as 0 = male, 1 = female), age, BMI, smoker (0 = smoker, 1 = non-smoker), verbal IQ, years of school education, and hours since awakening (e.g., Allen et al., 2017, 2014; Ginty et al., 2012; Kudielka et al., 2009, 2007; Lin et al., 2020; Slattery et al., 2013; Zänkert et al., 2019). Age, BMI, verbal IQ, years of school education, and hours since awakening were grand-mean centered. Covariates were not added together but considered in separate models. For more information on covariates, please refer to the Supplement. The time interval between TSST_1 and TSST_2 was not included in statistical models as time in months between TSST_1 and TSST_2 did not significantly correlate with change scores of TSST stress markers (Spearman: $rs = -.19$ to .26, $ps = .14$–.94).

2.4.3. Test-retest stability in areas-under-the-curve, absolute and relative peak change (TSST stress markers)

The same outliers that were excluded in the analysis of TSST trajectories were also excluded for the respective endocrinological measure in the analyses of TSST stress markers.

TSST stress markers were calculated for all outcome measures. Values measured right before the beginning of the TSST preparation period (T_2 for hormonal changes and T_1 for subjective stress) were used as the baseline response for calculation of TSST stress markers. This way, we tried to capture the stress response from right before the TSST until the end of the test session. Area-under-the-curve with respect to ground (AUC_G), as a measure of total hormone concentration and subjective stress, and area-under-the-curve with respect to increase (AUC_I), as a measure of change in hormonal concentrations and subjective stress, for variable time between measurements were calculated according to Pruessner et al. (2003). Absolute Peak Change (APC) was calculated by subtracting the baseline response from the maximum response. Relative Peak Change (RPC) was calculated as the percentage of change relative to the baseline response: APC/baseline response*100. For salivary cortisol, maximum response could be placed at all time-points from T_3 to T_7. For the other endocrinological measures, maximum response could be placed between T_3 and T_5, and for subjective stress either at T_2 or T_3.

AUC_G for salivary cortisol, ACTH, noradrenaline and subjective stress were log-transformed, and AUC_I, APC, and RPC for salivary and plasma cortisol, ACTH, noradrenaline and subjective stress were sqrt-transformed. Two-level LMMs, where individual TSST stress markers from the two test-days (level-1) were nested in individuals (level-2), with a random-intercept for participant ID were used to analyze differences in TSST stress markers between test-day (0 = TSST_1, 1 = TSST_2) and order (0 = early TSST, 1 = late TSST). To test the robustness of our results, we again conducted sensitivity analyses by including the above described covariates. As before, covariates were not added together, but considered in separate models.

To assess test-retest reliability of TSST stress markers in the subsample of individuals participating in both test-days, we correlated TSST stress markers from TSST_1 with the respective TSST stress marker from TSST_2 in Pearson product-moment correlation analyses. Scatterplots were used to visually inspect associations and to identify possible outliers driving correlations. One outlier was identified for salivary and plasma cortisol and excluded in sensitivity analyses. Furthermore, based on Miller et al. (2013), individuals were categorized as TSST responders or non-responders according to a relative increase in salivary cortisol of 15.5%. A chi-square test was used to analyze the responder non-responder ratios at TSST_1 and TSST_2.

2.4.4. General information

The significance level was set at $p \leq .050$. A false discovery rate (FDR) correction was used for the analyses of TSST stress markers (AUC_G, AUC_I, APC, RPC) within one outcome measure (Benjamini and Hochberg, 1995). All hierarchical data were analyzed with the 'nlme' package (Pinheiro et al., 2019) in R (R Core Team, 2019) and fitted with maximum likelihood estimation. Pseudo-R^2 was calculated according to Xu (2003) (1-(residual variance full model/residual variance null model)).

3. Results

3.1. Activation in the beginning of each test-day

LMMs indicated that salivary cortisol, plasma cortisol, ACTH, noradrenaline, and subjective stress levels at $T_{Beginning}$ were not significantly different between test-days ($ps > .14$). Thus, baseline differences at TSST_1 and TSST_2 were not observed.

3.2. TSST trajectories

3.2.1. Salivary cortisol

Repeated measurements of salivary cortisol within test-days (ICC(1) = 0.03, $F(1,509) = 8.68$, $p < .01$) and within individuals (ICC(1) = 0.49, $F(39,471) = 13.42$, $p < .001$) were non-independent. The ICC(2) for salivary cortisol samples within test-days was 0.88, and within individuals 0.93. Thus, the chosen level structure seems to be appropriate.

On average, salivary cortisol levels remained equal from the beginning of the test session (T_1; Table 2, Fig. 1) until after the TSST preparation phase at TSST_1 (T_3; preTSST slope). Salivary cortisol then significantly increased in response to the stress test until it reached its' peak 10 min after the TSST (T_5; reactivity slope: $b = +0.33$), followed by a significant decrease in salivary cortisol levels until 40 min after the TSST (T_7; recovery slope: $b = -0.13$). Regarding TSST_2, the interactions between time components and test-day did not become significant ($ps > .19$). Thus, the trajectory of the salivary cortisol response was not significantly changed at TSST_2. During TSST_1, salivary cortisol levels at T_1 were estimated lower when the TSST began late (late TSST: $b = -0.45$) due to the known circadian rhythm of cortisol secretion (Fig. S2).

Results remained robust in sensitivity analyses with additional covariates with the exception of order if hours since awakening was included. Order was no longer significant ($b = -0.12$, $p = .56$) whereas hours since awakening showed a significant effect ($b = -0.18$, $p = .01$). The longer individuals had been awake, the lower were salivary cortisol levels at T_1 during TSST_1.

Table 2
Discontinuous growth models for salivary and plasma cortisol, as well as plasma ACTH.

Salivary cortisol		Cortisol		ACTH	
Fixed effects	Coefficient (SE)	Fixed effects	Coefficient (SE)	Fixed effects	Coefficient (SE)
Intercept	1.24 (0.14)***	Intercept	95.49 (8.31)***	Intercept	3.90 (0.13)***
preTSST	0.00 (0.02)	preTSST	0.61 (2.25)	Reactivity	0.04 (0.01)***
Reactivity	0.33 (0.06)***	Reactivity 1	11.22 (2.81)***	Recovery 1	-0.08 (0.02)***
Recovery	-0.13 (0.02)***	Reactivity 2	-1.42 (2.66)	Recovery 2	-0.04 (0.02)
TSST_2	-0.14 (0.14)	Recovery	-4.96 (1.25)***	Recovery 3	0.02 (0.02)
TSST_2*preTSST	0.02 (0.03)	TSST_2	-1.58 (7.57)	TSST_2	-0.04 (0.06)
TSST_2*reactivity	-0.07 (0.06)	TSST_2*preTSST	-0.48 (1.86)	TSST_2*reactivity	0.01 (0.02)
TSST_2*recovery	-0.03 (0.03)	TSST_2*reactivity 1	2.14 (2.75)	TSST_2*recovery 1	-0.04 (0.03)
Late TSST	-0.45 (0.17)*	TSST_2*reactivity 2	-3.51 (2.27)	TSST_2*recovery 2	0.01 (0.04)
		TSST_2*recovery	-0.55 (1.10)	TSST_2*recovery 3	-0.01 (0.03)
Random effect variances	Estimate	Late TSST	-20.52 (10.22)	Late TSST	-0.13 (0.19)
Participant ID					
Intercept	0.40	Random effect variances	Estimate	Random effect variances	Estimate
preTSST	0.07	Participant ID		Participant ID	
Reactivity	0.30	Intercept	26.99	Intercept	0.55
Recovery	0.10	Test-day		Test-day	
Test-day		Intercept	18.87	Intercept	0.17
Intercept	0.55	Residual	24.47	Residual	0.21
preTSST	0.10				
Reactivity	0.24	BIC	4178.34	BIC	180.20
Recovery	0.07	R^2	17.85%	R^2	18.93%
Residual	0.19				
BIC	621.72				
R^2	87.09%				

Note. Time components indicate slopes per 10 min increments. Salivary cortisol: The analysis was based on seven measuring time-points on two test-days, TSST_1 and TSST_2, each; based on 40 individuals at TSST_1 and 33 individuals at TSST_2. Plasma cortisol and ACTH: The analysis was based on six measuring time-points on two test-days, TSST_1 and TSST_2, each; based on 39 individuals at TSST_1 and 32 individuals at TSST_2.
* $p < 0.05$; *** $p < 0.001$.

3.2.2. Plasma cortisol

Repeated assessments of plasma cortisol within test-days (ICC(1) = 0.02, $F(1,424) = 5.10$, $p = .03$) and within individuals (ICC(1) = 0.54, $F(38,387) = 13.74$, $p < .001$) were non-independent. The ICC(2) for plasma cortisol samples within test-days was 0.80, and within individuals 0.93.

Individuals experienced an increase in cortisol in reaction to the TSST at TSST_1 (T_3; reactivity 1 slope: $b = +11.22$; Table 2, Fig. 1) and cortisol levels stayed elevated until 20 min after the TSST (T_4; reactivity 2 slope). This was followed by a significant decrease in cortisol levels until the end of the test session (T_6; recovery slope: $b = -4.96$). The interactions between time components and test-day were not significant ($ps > .34$). By looking at the descriptive data (Fig. 1), we originally assumed that we would find a significant interaction between test-day*reactivity 2, leading to a faster recovery in cortisol levels at TSST_2. However, this was not the case ($p = .34$). Thus, as for salivary cortisol, TSST_2 did not significantly change the trajectory of the plasma cortisol response over time. Order was narrowly not significant (late TSST: $b = -20.52$, $p = .052$). Thus, during TSST_1, cortisol levels at T_1 were not significantly estimated lower when the TSST began late (Fig. S2).

However, if verbal IQ was included in sensitivity analyses, order became significant ($b = -20.89$, $p = .04$). In contrast, results remained robust if sex, age, BMI, smoker, years of education, or hours since awakening were added to the models. As we did not log-transform plasma cortisol values prior to analysis, we conducted sensitivity analyses with log-transformed cortisol levels which showed the same results as for non-log-transformed cortisol values. The only difference was a significant effect of order ($b = -0.29$, $p = .03$), with lower cortisol levels at T_1 during TSST_1 when the TSST began late. However, residuals for the model with log-transformed values did not approach normal distribution as well as for the model with non-log-transformed values.

3.2.3. ACTH

Repeated measurements of ACTH were non-independent within individuals (ICC(1) = 0.83, $F(38,387) = 54.16$, $p < .001$). This was not the case for test-days (ICC(1) = 0.00, $F(1,424) = 0.91$, $p = .34$). However, as we wanted to keep analyses similar between outcome measures, we nevertheless included test-day as a random effect. The ICC(2) for ACTH samples within test-days was − 0.10, and within individuals 0.98.

ACTH levels increased from the beginning of the test session and continued to increase in response to the TSST at TSST_1 (T_3; reactivity slope: $b = +0.04$; Table 2, Fig. 1). This was followed by an immediate decrease in ACTH levels until 20 min after the TSST (T_4; recovery 1 slope: $b = -0.08$). From there on, ACTH remained at the same level until the end of the test session. As for cortisol measures, the interactions between time components and test-day were not significant ($ps > .22$). Therefore, TSST_2 did not significantly change the trajectory of the ACTH response.

Moreover, order was not significant (late TSST: $b = -0.13$, $p = .49$), indicating that ACTH levels at T_1 during TSST_1 were the same for early and late onset of the TSST.

Results did not change in sensitivity analyses.

3.2.4. Noradrenaline

Repeated assessments of noradrenaline were non-independent within individuals (ICC(1) = 0.64, $F(39,381) = 19.65$, $p < .001$). The same was not true for test-days (ICC(1) = 0.00, $F(1,419) = 0.01$, $p = .94$). However, as we tried to keep analyses similar between outcome measures, we nevertheless included test-day as a random effect. The ICC(2) for noradrenaline samples within test-days was − 153.81, and within individuals 0.95.

On average, noradrenaline levels increased in reaction to the TSST (T_3; reactivity slope: $b = +0.18$; Table 3, Fig. 1) at TSST_1 and decreased immediately until 20 min later (T_4; recovery 1 slope: $b = -0.23$). From this time-point onwards noradrenaline levels increased slightly until the end of the test session (T_6; recovery 2 slope: $b = 0.02$). As for the other endocrinological measures, interactions between time components and test-day were not significant ($ps > .26$), indicating that the trajectory of the noradrenaline response did not significantly change during TSST_2.

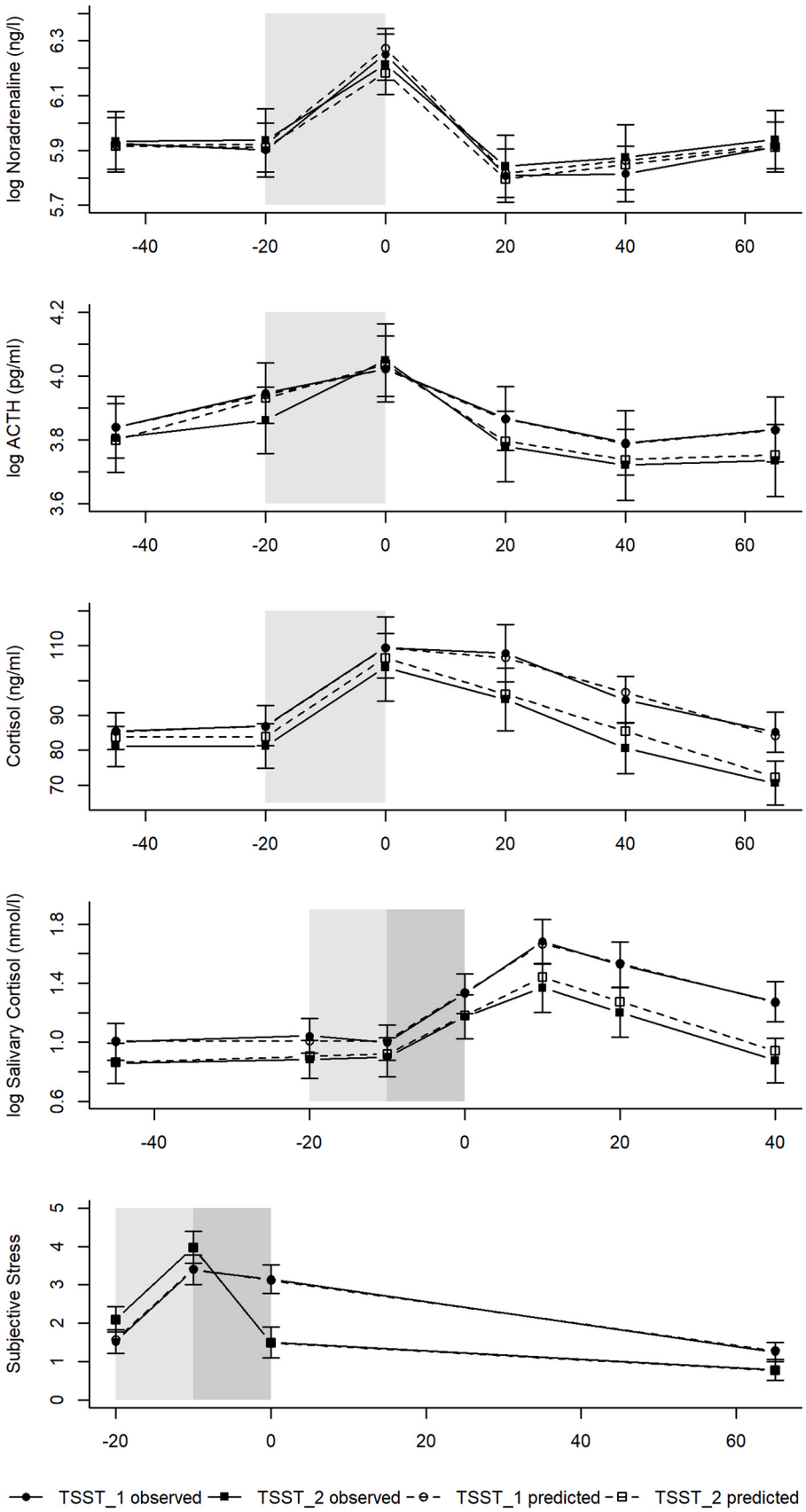

Fig. 1. Descriptive and predicted levels of plasma noradrenaline, plasma ACTH, plasma and salivary cortisol, and subjective stress over the course of the test sessions. Grey shaded areas indicate TSST periods.

Table 3
Discontinuous growth models for plasma noradrenaline and subjective stress.

Noradrenaline		Subjective stress	
Fixed effects	Coefficient (SE)	Fixed effects	Coefficient (SE)
Intercept	6.03 (0.11)***	Intercept	1.96 (0.38)***
preTSST	-0.00 (0.10)	TSST preparation	1.84 (0.36)***
Reactivity	0.18 (0.02)***	Reactivity	-0.29 (0.36)
Recovery 1	-0.23 (0.02)***	Recovery	-1.84 (0.35)***
Recovery 2	0.02 (0.01)*	TSST_2	0.52 (0.37)
TSST_2	-0.00 (0.10)	TSST_2*TSST preparation	0.04 (0.53)
TSST_2*preTSST	-0.00 (0.04)	TSST_2*reactivity	-2.19 (0.53)***
TSST_2*reactivity	-0.04 (0.04)	TSST_2*recovery	1.12 (0.53)*
TSST_2*recovery 1	0.04 (0.03)	Late TSST	-0.79 (0.44)
TSST_2*recovery 2	0.00 (0.02)		
Late TSST	-0.22 (0.15)	Random effect variances	Estimate
		Participant ID	
Random effect variances	Estimate	Intercept	1.27
Participant ID		Test-day	
Intercept	0.37	Intercept	0.00
Test-day		Residual	1.59
Intercept	0.36		
Residual	0.21	BIC	1275.29
		R^2	34.05%
BIC	200.73		
R^2	34.54%		

Note. Time components indicate slopes per 10 min increments. Noradrenaline: The analysis was based on six measuring time-points on two test-days, TSST_1 and TSST_2, each; based on 40 individuals at TSST_1 and 33 individuals at TSST_2; 17 noradrenaline measuring time-points were missing (see Supplement). Subjective stress: The analysis was based on four subjective stress ratings on two test-days, TSST_1 and TSST_2, each; based on 42 individuals at TSST_1 and 34 individuals at TSST_2; two subjects had missing data for one subjective stress rating during TSST_1.
* $p < 0.05$; *** $p < 0.001$.

During TSST_1, noradrenaline levels at T_1 were not estimated differently when the TSST began late (late TSST: $b = -0.22$, $p = .14$; Fig. S2).

Results remained robust against the inclusion of additional covariates.

To further show that endocrinological stress measures remained stable at TSST_2, we normalized all values by subtracting the stress levels measured at T_1 (− 45 min) from them. TSST trajectories with normalized values can be seen in Fig. S3.

3.2.5. Subjective stress

Repeated measurements of subjective stress were non-independent within individuals (ICC(1) = 0.31, $F(41,260) = 4.16$, $p < .001$). This was not the case for test-days (ICC(1) = 0.00, $F(1,300) = 0.86$, $p = .36$). However, as we wanted to keep analyses similar between outcome measures, we included test-day as a random effect. The ICC(2) for subjective stress within test-days was − 0.16, and within individuals 0.76.

At TSST_1, subjective stress was rated higher directly after the preparation period (T_2; TSST preparation slope: $b = +1.84$; Table 3, Fig. 1) and stayed elevated at around the same level until directly after the test period (T_3; reactivity slope), followed by a significant decrease until the end of the test session (T_4; recovery slope: $b = -1.84$). Contrary to the endocrinological measures, TSST_2 significantly changed the subjective feeling of stress. The TSST_2*TSST preparation interaction was not significant (TSST_2*TSST preparation: $b = 0.04$, $p = .94$), indicating that subjective stress ratings followed the same increase after the preparation period at TSST_2. However, the interactions between test-day and reactivity (TSST_2*reactivity: $b = -2.19$) as well as test-day and recovery (TSST_2*recovery: $b = 1.12$) were significant. Subjective stress ratings significantly fell from right before the test period until directly afterwards during TSST_2. Subsequently, subjective stress followed a less steep decrease in ratings until the end of the test session. This points to a faster recovery in the subjective experience of stress at TSST_2.

Order was not significant (late TSST: $b = -0.79$, $p = .08$). Subjective stress ratings before the TSST were thus not estimated differently when the TSST began late (Fig. S4).

In general, results remained robust in sensitivity analyses. Order became significant if age was included as an additional covariate ($b = -0.85$, $p = .04$). During TSST_1, subjective stress ratings at T_1 were estimated lower when the TSST began late if age was taken into account.

3.3. TSST stress markers – AUC_G, AUC_I, APC, RPC

As for TSST trajectories, AUC_G, AUC_I, APC and RPC remained similar for endocrinological measures at TSST_2 (*p*s > .15). Contrary to the TSST trajectories, late onset of the TSST did not significantly influence either of the TSST stress markers for endocrinological measures (*p*s > .39). The models are presented in Table S2. The inclusion of the covariates sex, age, BMI, smoker, verbal IQ, years of education, or hours since awakening in sensitivity analyses did not change the results.

As for the subjective stress trajectory, subjective stress AUC_G ($b = -70.22$, $p < .001$) and AUC_I ($b = -8.83$, $p < .001$) were smaller at TSST_2. However, APC ($p = .62$) and RPC ($p = .16$) did not significantly

Table 4
Test-retest-reliability between TSST_1 and TSST_2 for AUC_G, AUC_I, absolute, and relative peak change for individuals that participated in both test-days.

	Beginning	AUC_G	AUC_I	Absolute Peak Change	Relative Peak Change
Salivary cortisol ($n = 33$)	.131	.640***	.371*	.425*	.548**
Plasma cortisol ($n = 32$)	.192	.642***	.540***	.310	.363*
Plasma ACTH ($n = 32$)	.847***	.858***	-.176	-.146	-.204
Plasma noradrenaline ($n = 30$)	.535**	.606***	.456*	.475**	.338
Subjective stress ($n = 34$)	.450**	.837***	.294	.469**	.510**

Note. Pearson's product-moment correlation.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

change at TSST_2. Moreover, subjective stress markers were not influenced by later TSST onset (*ps* > .10). Sensitivity analyses showed that results remained robust against the inclusion of covariates.

3.4. Test-retest reliability

In the subsample of individuals participating in both test-days, test-retest reliability for all TSST stress markers as well as the value measured in the beginning of each test-day ranged from unacceptable to good. AUC_G showed the best test-retest reliability over all endocrinological and subjective measures (Table 4). For ACTH and subjective stress, reliability was good ($r > .837$, $p < .001$), whereas for both cortisol measures and noradrenaline reliability was only moderate ($r > .606$, $p < .001$). We had identified one outlier in salivary and plasma cortisol after visual inspection of scatterplots. Thus, to test the robustness of the found associations, we repeated the correlation analyses for salivary and plasma cortisol after exclusion of the outlier. Regarding salivary cortisol, the positive association between AUC_G at TSST_1 and TSST_2 remained ($r = .596$, $p < .001$, $n = 32$). The same was true for RPC ($r = .413$, $p = .02$, $n = 32$) whereas associations for AUC_I ($r = .229$, $p = .21$, $n = 32$) and APC ($r = .292$, $p = .11$, $n = 32$) were not maintained. With regard to plasma cortisol, only the association for RPC ($r = .275$, $p = .14$, $n = 31$) did not uphold, but all other associations remained largely unchanged (AUC_G: $r = .564$, $p < .001$, $n = 31$; AUC_I: $r = .465$, $p < .01$, $n = 31$; APC: $r = .221$, $p = .24$, $n = 31$).

During TSST_1, 24 individuals were responders and 9 were non-responders. This ratio remained the same during TSST_2 ($\chi^2(1) = 0.00$, $p = 1.00$). Importantly, this was not based on the same individuals. 16 individuals changed their responder-group at TSST_2 with 8 individuals becoming non-responders and 8 actually becoming responders.

4. Discussion

The goal of this study was to test the stability of several HPA and SAM axis biomarkers as well as subjective stress responses after two exposures to the TSST four months apart. This is the first longitudinal TSST investigation combining a large number of physiological stress markers and applying discontinuous growth modelling of the response curves. Summarizing the main results, we did not observe a decrease in the response magnitude of the endocrinological biomarkers nor of the subjective stress response at TSST_2. AUC_G was the most reliable measure across all outcome measures.

As hypothesized the course of the TSST response was similar between TSST_1 and TSST_2 in all endocrinological outcome measures. Thus, previously observed test-retest effects of the HPA axis at shorter follow-up intervals did not occur after a 4-month interval. This aligns with Petrowski et al. (2012) who did not observe salivary cortisol habituation after a 10-week interval. Remarkably, robust test-retest stability occurred in our study, even though the TSST protocol was kept constant between test-days. It has been proposed that slight changes to the TSST protocol, such as changing the speech topic and arithmetic task, might be done to avoid test-retest effects between two TSST exposures (Kirschbaum et al., 1995). However, longitudinal studies usually require constant settings across all measurement points, wherefore we decided to not make any changes on the second TSST. Considering our results, it seems that the longer time interval might already suffice to reinstate the original stress response. Descriptively, salivary and plasma cortisol levels were slightly lower at TSST_2 (Fig. 1). This is also reflected, again only at a descriptive level, in a smaller salivary cortisol AUC_G at TSST_2. As AUC_G is seen as a measure of total hormone concentration (Pruessner et al., 2003), this might suggest a weak and non-significantly lower cortisol output in saliva during TSST_2. It is conceivable that individuals had gotten slightly used to the study's general environment and that overall anticipatory stress levels were somewhat decreased at TSST_2. As in previous studies (Boesch et al., 2014; Gerra et al., 2001; Schommer et al., 2003), we did not observe test-retest effects of the SAM axis, as measured by noradrenaline, over repeated exposure. Regarding the subjective stress response, test-retest effects occurred as indicated by the TSST trajectory, AUC_G, and AUC_I, but not APC and RPC. Overall, this reflects a faster recovery after the second TSST rather than a diminished subjective stress response as the increase in stress after the TSST preparation period was estimated to be the same during both TSSTs. Whereas the subjective stress level kept elevated until right after the TSST during the first exposure, subjective stress levels dropped directly after the second. We therefore assume that individuals entered the test period with the same subjective stress level during both test-days but felt a faster relief directly after the second TSST. This shows the importance of assessing subjective stress or any other subjective outcome at multiple time-points during the TSST. Simple pre-post comparisons of subjective outcomes, which are generally done in TSST habituation studies, may have overlooked this finding. Schlotz et al. (2008) and Campbell and Ehlert (2012) already proposed previously that repeated assessments of emotional states are better suited to display the rapidly changing subjective stress experience. Moreover, due to different periods, the TSST is a task that is particularly suited to investigate time-dependent effects on the subjective stress response. Sensitivity analyses showed that results regarding the influence of test-day were robust against the inclusion of covariates such as sex, age, BMI, smoking, verbal IQ, and hours since awakening. Altogether, the TSST seems to be suitable for longitudinal studies that require stable stress responses.

AUC_G showed the best associations for all outcome measures. Strong associations were found for ACTH and subjective stress ($r \geq .837$) and moderate associations were found for the cortisol measures and noradrenaline ($r \geq .606$). Thus, for AUC_G test-retest reliability was especially good for ACTH and subjective stress. Regarding cortisol measures and noradrenaline, test-retest reliability was still acceptable but became hardly acceptable for cortisol measures once the outlier was excluded. Test-retest reliability for TSST stress markers indicating change (AUC_I, APC, RPC) in hormonal concentrations or subjective stress was in general moderate to poor. The majority of individuals remained responders at TSST_2. However, some individuals changed their responder-group experiencing either habituation or sensitization. The latter could be expected as previous studies have shown that a minority of individuals experiences sensitization over repeated exposure (Kudielka et al., 2006; Wüst et al., 2005). Nevertheless, at a group-level, we did not observe strong test-retest effects in our sample. However, the responder analysis as well as the heterogenous test-retest reliability ranging from poor to moderate for TSST stress markers indicating change show that the data analysis is likely constrained in the detection of individual changes in the response pattern in a longitudinal design. Thus, it has to be kept in mind that individual changes may occur over the course of 4-months but that at a group-level the response pattern remained stable between two TSSTs four months apart.

As expected, salivary cortisol followed the known circadian rhythm (Debono et al., 2009; Krieger et al., 1971; Weitzman et al., 1971), with higher cortisol levels when the TSST was conducted early and lower cortisol levels when the TSST was conducted later in the afternoon. Surprisingly, AUC_G, that summarizes total hormone concentration, did not reflect this finding. The other salivary cortisol TSST stress markers were also not affected by TSST order. All other outcome measures were not significantly affected by order. However, results could change if covariates were included.

The results have to be interpreted with the following limitations in mind. The sample size was small. Therefore, it is possible that the power to detect differences between test-days was not sufficient. One reason for the moderate power of the longitudinal analysis was that some participants dropped out as they did not return for the TSST_2 measurement. However, as LMMs are good in handling missing data (DeShon et al., 1998; Laird, 1988; Pinheiro and Bates, 2000), we preferred to use the entire sample to not lose important data. Moreover, our sample had quite a large age range as well as a rather uneven gender distribution. Furthermore, due to organizational constraints, we did not always

succeed in testing women during the luteal phase of their menstrual cycle. Thus, results should be replicated in a larger and more homogenous sample.

A 4-month interval is a sufficient time interval between two repeated TSST exposures to fully reinstate the initial physiological and psychological stress response. This was true although the TSST protocol was not changed. TSST trajectories and TSST stress markers were largely in accordance regarding each outcome measure. The trajectory allows a more fine-grained analysis of the stress response whereas the TSST stress markers either give an indication of general output (AUC_G) or change (AUC_I, APC, RPC). Thus, to fully grasp the stress response, we recommend to always analyze the detailed TSST trajectory as well as the more compressed TSST stress markers if the stress response is the main outcome. This would also facilitate comparison between studies. AUC_G was the most reliable marker for all outcome measures, whereas TSST stress markers indicating change were much less reliable. Concluding, the TSST seems to be suitable for the application in longitudinal studies that require a stable stress response at a group-level. However, individual changes in the response pattern may occur (such as change in responder status), which have to be considered in the interpretation of such data.

Funding

The study was supported by a grant from the Swiss National Science Foundation, Switzerland (Grant No. 105319_162639), to BBQ. BKS received a grant from the Coordination for the Improvement of Higher Education Personnel, CAPES, Brazil (Grant No. 99999.001968/2015-07). The funders had no role in study design; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript, and the decision to submit the manuscript for publication.

CRediT authorship contribution statement

BBQ developed the study concept/design and rose the funding for the study. AKK, BKS, and MV conducted the assessments. LMS helped with the implementation of the TSST, trained the experimenters in the related protocol, and supported the data interpretation. CK conducted the saliva and plasma analyses and supported the data interpretation. AKK conducted the statistical analyses. AKK and BBQ drafted the first manuscript. All authors contributed to and approved of the final manuscript.

Declaration of interest

None.

Acknowledgements

We are grateful to Monika Näf, Chantal Kunz, Marlon Nüscheler, Selina Maisch, Jocelyn Waser, Anna Burkert, Meret Speich, Maxine de Ven, Zoé Dolder, Zoe Hillmann, Jessica Grub, and Priska Cavegn for their excellent support with recruitment and assessment of the participants.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psyneuen.2021.105342.

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.

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. Neurobiol. Stress 6, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001.

Asbrand, J., Heinrichs, N., Nitschke, K., Wolf, O.T., Schmidtendorf, S., Tuschen-Caffier, B., 2019. Repeated stress leads to enhanced cortisol stress response in child social anxiety disorder but this effect can be prevented with CBT. Psychoneuroendocrinology 109, 104352. https://doi.org/10.1016/j.psyneuen.2019.06.003.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.

Boesch, M., Sefidan, S., Ehlert, U., Annen, H., Wyss, T., Steptoe, A., La Marca, R., 2014. Mood and autonomic responses to repeated exposure to the Trier Social Stress Test for Groups (TSST-G). Psychoneuroendocrinology 43, 41–51. https://doi.org/10.1016/j.psyneuen.2014.02.003.

Campbell, J., Ehlert, U., 2012. Acute psychosocial stress: Does the emotional stress response correspond with physiological responses? Psychoneuroendocrinology 37, 1111–1134. https://doi.org/10.1016/j.psyneuen.2011.12.010.

Chen, X., Gianferante, D., Hanlin, L., Fiksdal, A., Breines, J.G., Thoma, M.V., Rohleder, N., 2017. HPA-axis and inflammatory reactivity to acute stress is related with basal HPA-axis activity. Psychoneuroendocrinology 78, 168–176. https://doi.org/10.1016/j.psyneuen.2017.01.035.

Debono, M., Ghobadi, C., Rostami-Hodjegan, A., Huatan, H., Campbell, M.J., Newell-Price, J., Darzy, K., Merke, D.P., Arlt, W., Ross, R.J., 2009. Modified-release hydrocortisone to provide circadian cortisol profiles. J. Clin. Endocrinol. Metab. 94, 1548–1554. https://doi.org/10.1210/jc.2008-2380.

DeShon, R.P., Ployhart, R.E., Sacco, J.M., 1998. The estimation of reliability in longitudinal models. Int. J. Behav. Dev. 22, 493–515. https://doi.org/10.1080/016502598384243.

Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol. Bull. 130, 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, 116–121. https://doi.org/10.1037/0278-6133.27.1.116.

Federenko, I.S., Nagamine, M., Hellhammer, D.H., Wadhwa, P.D., Wüst, S., 2004. The heritability of hypothalamus pituitary adrenal axis responses to psychosocial stress is context dependent. J. Clin. Endocrinol. Metab. 89, 6244–6250. https://doi.org/10.1210/jc.2004-0981.

Foley, P., Kirschbaum, C., 2010. Human hypothalamus–pituitary–adrenal axis responses to acute psychosocial stress in laboratory settings. Neurosci. Biobehav. Rev. 35, 91–96. https://doi.org/10.1016/j.neubiorev.2010.01.010.

Gerra, G., Zaimovic, A., Mascetti, G.G., Gardini, S., Zambelli, U., Timpano, M., Raggi, M. A., Brambilla, F., 2001. Neuroendocrine responses to experimentally-induced psychological stress in healthy humans. Psychoneuroendocrinology 26, 91–107. https://doi.org/10.1016/S0306-4530(00)00046-9.

Gianferante, D., Thoma, M.V., Hanlin, L., Chen, X., Breines, J.G., Zoccola, P.M., Rohleder, N., 2014. Post-stress rumination predicts HPA axis responses to repeated acute stress. Psychoneuroendocrinology 49, 244–252. https://doi.org/10.1016/j.psyneuen.2014.07.021.

Ginty, A.T., Phillips, A.C., Roseboom, T.J., Carroll, D., DeRooij, S.R., 2012. Cardiovascular and cortisol reactions to acute psychological stress and cognitive ability in the Dutch Famine Birth Cohort Study. Psychophysiology 49, 391–400. https://doi.org/10.1111/j.1469-8986.2011.01316.x.

Gold, S.M., Zakowski, S.G., Valdimarsdottir, H.B., Bovbjerg, D.H., 2004. Higher Beck depression scores predict delayed epinephrine recovery after acute psychological stress independent of baseline levels of stress and mood. Biol. Psychol. 67, 261–273. https://doi.org/10.1016/j.biopsycho.2003.12.001.

Gonzalez-Bono, E., Rohleder, N., Hellhammer, D.H., Salvador, A., Kirschbaum, C., 2002. Glucose but not protein or fat load amplifies the cortisol response to psychosocial stress. Horm. Behav. 41, 328–333. https://doi.org/10.1006/hbeh.2002.1766.

Hellhammer, J., Schubert, M., 2012. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. Psychoneuroendocrinology 37, 119–124. https://doi.org/10.1016/j.psyneuen.2011.05.012.

Hernández-Lloreda, M.V., Colmenares, F., Martínez-Arias, R., 2004. Application of piecewise hierarchical linear growth modeling to the study of continuity in behavioral development of baboons (Papio hamadryas). J. Comp. Psychol. 118, 316–324. https://doi.org/10.1037/0735-7036.118.3.316.

Het, S., Vocks, S., Wolf, J.M., Herpertz, S., Wolf, O.T., 2020. Treatment-resistant blunted HPA activity, but reversible cardiovascular stress reactivity in young women with eating disorders. Front. Psychiatry 11, 1–10. https://doi.org/10.3389/fpsyt.2020.00726.

Höhne, N., Poidinger, M., Merz, F., Pfister, H., Brückl, T., Zimmermann, P., Uhr, M., Holsboer, F., Ising, M., 2014. Increased HPA axis response to psychosocial stress in remitted depression: the influence of coping style. Biol. Psychol. 103, 267–275. https://doi.org/10.1016/j.biopsycho.2014.09.008.

Jezova, D., Makatsori, A., Duncko, R., Moncek, F., Jakubek, M., 2004. High trait anxiety in healthy subjects is associated with low neuroendocrine activity during psychosocial stress. Prog Neuropsychopharmacol Biol Psychiatry 28, 1331–1336. https://doi.org/10.1016/j.pnpbp.2004.08.005.

Jones, P.R., 2019. A note on detecting statistical outliers in psychophysical data. Atten Percept Psychophys 81, 1189–1196. https://doi.org/10.3758/s13414-019-01726-3.

Jönsson, P., Wallergård, M., Österberg, K., Hansen, Å.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, 1397–1403. https://doi.org/10.1016/j.psyneuen.2010.04.003.

Jönsson, P., Österberg, K., Wallergård, M., Hansen, Å.M., Garde, A.H., Johansson, G., Karlson, B., 2015. Exhaustion-related changes in cardiovascular and cortisol reactivity to acute psychosocial stress. Physiol. Behav. 151, 327–337. https://doi.org/10.1016/j.physbeh.2015.07.020.

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, 76–81. https://doi.org/10.1159/000119004.

Kirschbaum, C., Prussner, J.C., Stone, A.A., Federenko, I., Gaab, J., Lintz, D., Schommer, N., Hellhammer, D.H., 1995. Persistent high cortisol responses to repeated psychological stress in a subpopulation of healthy men. Psychosom. Med. 57, 468–474. https://doi.org/10.1097/00006842-199509000-00009.

Kirschbaum, C., Gonzalez Bono, E., Rohleder, N., Gessner, C., Pirke, K.M., Salvador, A., Hellhammer, D.H., 1997. Effects of fasting and glucose load on free cortisol responses to stress and nicotine. J. Clin. Endocrinol. Metab. 82, 1101–1105. https://doi.org/10.1210/jc.82.4.1101.

Kirschbaum, C., Kudielka, B.M., Gaab, J., Schommer, N.C., Hellhammer, D.H., 1999. Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. Psychosom. Med. 61, 154–162. https://doi.org/10.1097/00006842-199903000-00006.

Krieger, D.T., Allen, W., Rizzo, F., Krieger, H.P., 1971. Characterization of the normal temporal pattern of plasma corticosteroid levels. J. Clin. Endocrinol. Metab. 32, 266–284. https://doi.org/10.1210/jcem-32-2-266.

Kudielka, B.M., von Känel, R., Preckel, D., Zgraggen, L., Mischler, K., Fischer, J.E., 2006. Exhaustion is associated with reduced habituation of free cortisol responses to repeated acute psychosocial stress. Biol. Psychol. 72, 147–153. https://doi.org/10.1016/j.biopsycho.2005.09.001.

Kudielka, B.M., Hellhammer, D.H., Kirschbaum, C., 2007. Ten years of research with the Trier Social Stress Test (TSST) - revisited. In: Harmon-Jones, E., Winkielman, P. (Eds.), Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior. Guilford Press, New York, NY, pp. 56–83.

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, 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. Neurosci. Biobehav. Rev. 107, 686–695. https://doi.org/10.1016/j.neubiorev.2019.09.032.

Laird, N.M., 1988. Missing data in longitudinal studies. Stat. Med. 7, 305–315. https://doi.org/10.1002/sim.4780070131.

Lin, L., Wu, J., Yuan, Y., Sun, X., Zhang, L., 2020. Working memory predicts hypothalamus-pituitary-adrenal axis response to psychosocial stress in males. Front. Psychiatry 11, 1–9. https://doi.org/10.3389/fpsyt.2020.00142.

Manigault, A.W., Zoccola, P.M., Wüst, S., Yim, I.S., 2020. Corroborative evidence for an association between initial hypothalamic-pituitary-adrenocortical axis reactivity and subsequent habituation in humans. Psychoneuroendocrinology 121, 104798. https://doi.org/10.1016/j.psyneuen.2020.104798.

McEwen, B.S., 1998. Stress, adaptation, and disease: allostasis and allostatic load. Ann. N. Y. Acad. Sci. 840, 33–44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x.

McEwen, B.S., Stellar, E., 1993. Stress and the individual. Mechanisms leading to disease. Arch. Intern. Med. 153, 2093–2101.

McInnis, C.M., Wang, D., Gianferante, D., Hanlin, L., Chen, X., Thoma, M.V., Rohleder, N., 2015. Response and habituation of pro- and anti-inflammatory gene expression to repeated acute stress. Brain Behav Immun 46, 237–248. https://doi.org/10.1016/j.bbi.2015.02.006.

Miller, R., Plessow, F., Kirschbaum, C., Stalder, T., 2013. Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: evaluation of salivary cortisol pulse detection in panel designs. Psychosom. Med. 75, 832–840. https://doi.org/10.1097/PSY.0000000000000002.

Petrowski, K., Herold, U., Joraschky, P., Wittchen, H.-U., Kirschbaum, C., 2010. A striking pattern of cortisol non-responsiveness to psychosocial stress in patients with panic disorder with concurrent normal cortisol awakening responses. Psychoneuroendocrinology 35, 414–421. https://doi.org/10.1016/j.psyneuen.2009.08.003.

Petrowski, K., Wintermann, G.-B., Siepmann, M., 2012. Cortisol response to repeated psychosocial stress. Appl. Psychophysiol. Biofeedback 37, 103–107. https://doi.org/10.1007/s10484-012-9183-4.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2019. nlme: linear and nonlinear mixed effects models, R package version 3.1-142.

Pinheiro, J.C., Bates, D.C., 2000. Mixed-Effects Models in S and S-PLUS, Statistics and Computing. Springer Publishing Company, New York, NY. https://doi.org/10.1007/b98882.

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, 916–931. https://doi.org/10.1016/S0306-4530(02)00108-7.

Quednow, B.B., Kühn, K.U., Hoenig, K., Maier, W., Wagner, M., 2004. Prepulse inhibition and habituation of acoustic startle response in male MDMA ('Ecstasy') users, cannabis users, and healthy controls. Neuropsychopharmacology 29, 982–990. https://doi.org/10.1038/sj.npp.1300396.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.r-project.org/.

Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical linear models: applications and data analysis methods, 2nd ed. Sage, Thousand Oaks, CA.

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? Ann. N. Y. Acad. Sci. 1032, 258–263. https://doi.org/10.1196/annals.1314.033.

Roos, L.G., Janson, J., Sturmbauer, S.C., Bennett, J.M., Rohleder, N., 2019. Higher trait reappraisal predicts stronger HPA axis habituation to repeated stress. Psychoneuroendocrinology 101, 12–18. https://doi.org/10.1016/j.psyneuen.2018.10.018.

Schlotz, W., Kumsta, R., Layes, I., Entringer, S., Jones, A., Wüst, S., 2008. Covariance between psychological and endocrine responses to pharmacological challenge and psychosocial stress: a question of timing. Psychosom. Med. 70, 787–796. https://doi.org/10.1097/PSY.0b013e3181810658.

Schommer, N.C., Hellhammer, D.H., Kirschbaum, C., 2003. Dissociation between reactivity of the hypothalamus-pituitary-adrenal axis and the sympathetic-adrenal-medullary system to repeated psychosocial stress. Psychosom. Med. 65, 450–460. https://doi.org/10.1097/01.PSY.0000035721.12441.17.

Singer, J.D., Willett, J.B., 2003. Applied longitudinal data analysis. Oxford University Press, Oxford, UK.

Slattery, M.J., Grieve, A.J., Ames, M.E., Armstrong, J.M., Essex, M.J., 2013. Neurocognitive function and state cognitive stress appraisal predict cortisol reactivity to an acute psychosocial stressor in adolescents. Psychoneuroendocrinology 38, 1318–1327. https://doi.org/10.1016/j.psyneuen.2012.11.017.

Sugaya, N., Izawa, S., Kimura, K., Ogawa, N., Yamada, K.C., Shirotsuki, K., Mikami, I., Hirata, K., Nagano, Y., Nomura, S., Shimada, H., 2012. Adrenal hormone response and psychophysiological correlates under psychosocial stress in individuals with irritable bowel syndrome. Int. J. Psychophysiol. 84, 39–44. https://doi.org/10.1016/j.ijpsycho.2012.01.006.

Thoma, M.V., Kirschbaum, C., Wolf, J.M., Rohleder, N., 2012. Acute stress responses in salivary alpha-amylase predict increases of plasma norepinephrine. Biol. Psychol. 91, 342–348. https://doi.org/10.1016/j.biopsycho.2012.07.008.

Weitzman, E.D., Fukushima, D., Nogeire, C., Roffwarg, H., Gallagher, T.F., Hellman, L., 1971. Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. J. Clin. Endocrinol. Metab. 33, 14–22. https://doi.org/10.1210/jcem-33-1-14.

Wittchen, H.U., Wunderliche, U., Gruschwitz, S., Zaudig, M., 1997. SKID-I. Strukturiertes Klinisches Interview für DSM-IV Achse I: Psychische Störungen [SCID-I. Structured Clinical Interview for DSM-IV Axis I: Mental Disorders]. Hogrefe, Göttingen.

Wolf, O.T., Schommer, N.C., Hellhammer, D.H., Reischies, F.M., Kirschbaum, C., 2002. Moderate psychosocial stress appears not to impair recall of words learned 4 weeks prior to stress exposure. Stress 5, 59–64. https://doi.org/10.1080/1025389029000123 32.

Wüst, S., Federenko, I.S., van Rossum, E.F.C., Koper, J.W., Hellhammer, D.H., 2005. Habituation of cortisol responses to repeated psychosocial stress—further characterization and impact of genetic factors. Psychoneuroendocrinology 30, 199–211. https://doi.org/10.1016/j.psyneuen.2004.07.002.

Xu, R., 2003. Measuring explained variation in linear mixed effects models. Stat. Med. 22, 3527–3541. https://doi.org/10.1002/sim.1572.

Zänkert, S., Bellingrath, S., Wüst, S., Kudielka, B.M., 2019. HPA axis responses to psychological challenge linking stress and disease: what do we know on sources of intra- and interindividual variability? Psychoneuroendocrinology 105, 86–97. https://doi.org/10.1016/j.psyneuen.2018.10.027.