Computers in Human Behavior Reports 16 (2024) 100497

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Validation of the Virtual Reality Stroop Room: Effects of inhibiting interfering information under time-pressure and task-switching demands

Veronika Ringgold [a,b,*], Luca Abel [b], Bjoern M. Eskofier [b,c], Nicolas Rohleder [a]

[a] Chair of Health Psychology, Department of Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[b] Machine Learning and Data Analytics Lab, Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[c] Institute of AI for Health, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

ARTICLE INFO

Keywords:
Virtual reality
Stress
Stroop
Executive functions
Inhibition

ABSTRACT

The physiological stress response affects executive functions, such as inhibition, as assessed by the Stroop Color and Word Test. In this study, we investigated the effects of the Virtual Reality Stroop Room (VRSR), a research paradigm assessing these cognitive top-down processes while inducing mild acute stress, on self-reported stress states, heart rate, salivary alpha-amylase, and cortisol. Our sample consisted of 89 participants (52 women; Age: 23.60 ± 3.88 years) and was evenly allocated to the three conditions of the VRSR (*regular, time pressure,* and *rotation*). The Stroop Effect, reflected in prolonged processing times and increased errors in the incongruent phase, was observed. Participants reported heightened *Distress* and *Engagement* post-experiment, alongside lower *Worry*, assessed via the Short Stress State Questionnaire. Scores from the Positive and Negative Affect Schedule indicated elevated positive affect and decreased negative affect post-study. With regard to biosignals we found that heart rate was higher in the incongruent phase, compared to the congruent phase and a significant time × condition interaction was observed. Salivary alpha-amylase exhibited a significant time effect. Results for cortisol do not support a uniform response of the hypothalamus-pituitary-adrenal (HPA) axis. In conclusion, the VRSR appears to be a valid measure for executive functions while activating the sympathetic nervous system, but not the HPA axis. Its current implementation induces mild physiological and psychological stress responses, with fewer adverse reactions compared to the Trier Social Stress Test. Future studies should leverage the adaptability of virtual reality applications to refine this research paradigm.

1. Introduction

At the time this article was written, a search for articles concerning *stress* and *health* on PubMed resulted in 337,350 articles. That same search on Google Scholar resulted in over 5.7 million articles with over 50,000 alone since 2023. What is more, with costs between approximately 221 million and 187 billion US$ in 2014 for work-stress-related mental health outcomes (Hassard et al., 2018), the importance of research concerning stress cannot be overstated. There is no question that stress is among the leading contributors to negative physiological and psychological health outcomes (Cohen et al., 2007; Rohleder, 2012; Schneiderman et al., 2005; Shields & Slavich, 2017). While it is undisputed that it is chronic stress that negatively impacts health, the transition from acute to chronic stress is still not entirely understood (Rohleder, 2019; Shields & Slavich, 2017). Early conceptualizations have focused on universal stressors which were thought to affect most people (Holmes & Rahe, 1967). A more widely accepted understanding today is that an individual's perception and appraisal of a potentially stressful situation shape their psychological and physiological stress response (Lazarus & Folkman, 1984). From a biological perspective, wear-and-tear processes are conceptualized as the mechanisms through which stress turns into physiological damage (McEwen, 1998). However, the processes that connect these two constructs are not understood to an extent which would make further explanatory variables redundant. While cognition has long been a factor of interest (Shields, 2020), a more recent model of how stress and psychopathology might be connected through cognition is the *Integrated Model of Stress, Executive Control, and Psychopathology* (Quinn & Shields, 2023). This model proposes a pathway from stress to psychopathology through deteriorating executive control, with results such as more rumination or less effective

---

* Corresponding author. Chair of Health Psychology, Department of Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
E-mail address: veronika.ringgold@fau.de (V. Ringgold).

https://doi.org/10.1016/j.chbr.2024.100497
Received 10 June 2024; Accepted 25 September 2024

cognitive reappraisal. The model explains biological, emotional and cognitive differences in individuals' responses to stress with regard to cognitive functioning and posits that to connect stress and psychopathology through executive control, differences must exist in the effects of stress on executive control between individuals. From this it follows that research is in dire need of tools to explore these differences. The gold standard protocol to elicit acute stress, the Trier Social Stress Test (TSST; Kirschbaum et al., 1993) uses stress-inducing factors such as social evaluation and a demanding mental arithmetic task. However, neither the TSST nor, to our knowledge, any other acute stress protocol, makes use of higher-level executive functions (EF) during acute stress while at the same time allowing for their quantification.

When it comes to stress, researchers have rightly been criticized for applying the construct too broadly and using single indicators, which cannot capture this complex phenomenon (Kagan, 2016; McEwen & McEwen, 2016). Therefore, it is important to not only measure acute stress responses repeatedly (Rohleder, 2019), but to assess them via psychological and physiological stress markers. Two stress systems influence how the body reacts to a potentially stressful situation. One, the autonomous nervous system (ANS) – with its two antagonists, the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) – and two, the hypothalamus-pituitary-adrenal (HPA) axis. The ANS is responsible for the so-called fight-or-flight response by quickly activating bodily resources, such as increasing the heart rate, or releasing epinephrine and norepinephrine into the bloodstream. The activation of the SNS can be measured using biosignals, such as the electrocardiogram (ECG), or biomarkers, such as salivary alpha-amylase (sAA). The slower reacting stress response system, the HPA axis, secretes various glucocorticoids, such as cortisol. Cortisol serves an important role in bodily changes and can affect nearly every organ system, including the immune, cardiovascular, and nervous systems (Hellhammer et al., 2009; Kuras et al., 2017; Strahler et al., 2017). It serves as a regulatory system and, bridging the conceptual gap between stress and EF, is influenced by a person's cognitions and reactions to a situation (Miller et al., 2007).

As with stress, there is wide agreement on an overall taxonomy of EF or executive control (Diamond, 2013; Miyake & Friedman, 2012). Yet, there are differences when it comes to the details of how the components may be structured or how they affect peoples' daily life. Summarizing the various EF definitions in a slightly simplified statement, they are a set of cognitive processes that allow an individual to choose how to respond to their surrounding circumstances, instead of mainly reacting to them. Depending on the model, EF are either seen as a general component that is also influenced by the specific abilities *shifting* and *updating* (unity-diversity-framework; Miyake et al., 2000; Miyake & Friedman, 2012), or as a set of abilities that is comprised of *inhibition, working memory* and *cognitive flexibility* (Diamond, 2013). In either conceptualization, EF are the building blocks for higher level cognitive functions, such as planning, problem solving or reasoning, which are also sometimes subsumed under the term fluid intelligence (Diamond, 2013). It is important to note though that EF are not the same as intelligence or the intelligence quotient (IQ), even though they are strongly connected (Miyake & Friedman, 2012). For the scope of this paper, we will use the terms inhibition, updating – which corresponds roughly to processes surrounding working memory –, and shifting – which is approximately equivalent to cognitive flexibility (Diamond, 2013; Miyake & Friedman, 2012). Giving a more detailed overview over the different conceptualizations and models, lies outside the scope of this paper and can be found elsewhere (Karr et al., 2018; Quinn & Shields, 2023).

A link between EF and stress is well established (Quinn & Shields, 2023; Shields, 2020; Shields et al., 2015, 2016), although the connection does not seem to be a simple one. Some facets of EF are enhanced by stress – mainly those of motor actions –, while some, such as the executive control of thoughts, appear on the whole to be impaired (Shields et al., 2016). With stressful life experiences being a highly relevant contributing factor with regard to developing psychopathology, executive functions – with their link to stress – need to be considered when examining this connection (Grant et al., 2004; Quinn & Shields, 2023; Snyder et al., 2015). To date, it is unclear how deficits in executive control and its facets are related to a maladaptive stress response and, consequently, over time, to psychopathology (Quinn & Shields, 2023). Thus, research aiming at understanding stress, psychopathology/health and EF in conjunction is in need of approaches examining executive control under stress, to understand how stress, through EF, might influence psychopathology (Quinn & Shields, 2023; Shields, 2020).

One well known research paradigm investigating EF is the Stroop Color and Word Test (SCWT; Stroop, 1935). In its basic implementation the SCWT uses color words which are shown either in the color the word spells out, e.g., the word "Blue" written in blue ink (congruent), or displaying a mismatch between these two dimensions, e.g., the word "Blue" written in red ink (incongruent). Respondents are asked to name the color of the ink, which in the congruent version of the test is very simple. Yet, in the incongruent version the prepotent response of the brain is to process the word, thus creating interference when trying to name the color the word is written in. This cognitive interference leads to longer reaction times and more errors in the incongruent phase of the SCWT, famously named the *Stroop Effect* after the developer of the test (Stroop, 1935). While this effect has been replicated many times (MacLeod, 1991; Scarpina & Tagini, 2017), some researchers also found that the SCWT results in a physiological stress response (Fauvel et al., 1996; Renaud & Blondin, 1997; Tulen et al., 1989). Again other results show that stress influences performance in the SCWT (Henderson et al., 2012). Considering the theoretical framework outlined by Quinn & Shields, 2023, our study aims to investigate whether the SCWT elicits a physiological stress response or arousal, and if so, whether it can serve as a reliable measure of executive functions under acute stress.

To investigate this, we make use of the Virtual Reality Stroop Room (VRSR; Gradl et al., 2019), a 3D virtual reality (VR) implementation of the Stroop Test with an added task-switching component, the possibility to alter the sequence of the congruent and incongruent phase, and additional features such as the option to change the time per trial or rotating the walls showing the colors. While the SCWT strongly taps into the inhibition component of EF, the VRSR, in the incongruent phase, requires the use of all three EF components. Participants have to shift between tasks, either reacting to the word or color, and subsequently in each trial have to inhibit either the word or the color of the color-word. Further, they have to constantly update their spatial location in the VRSR to find the correct wall as quickly as possible. This has the potential to make the VRSR a highly useful tool in psychological research, as it allows for high experimental control and relatively quick implementation of changes or additional factors of interest, such as sounds or visual stimuli.

The aim of the present study is to evaluate the overall effects of the VRSR on physiological and psychological outcomes regarding arousal, stress, and affect. Further, we want to provide researchers with information about the executive function measure of the VRSR by analyzing errors and reaction times in both phases (Scarpina & Tagini, 2017). The questions we aim to answer are if the VRSR can reliably replicate the Stroop Effect and if it leads to a physiological stress response, thus allowing it to be used for research at the intersection of EF and stress.

2. Methods and materials

During the experiment we logged participants' performance data, namely task completion times and errors, continuously recorded their ECG, and took five saliva samples over the course of the experiment to assess their sAA and cortisol levels. Further, participants were asked to fill out self-report measures of affect and stress before and after the experiment. The study was designed as a stratified randomized parallel-group study with three between factors (*regular:* Baseline implementation of the VRSR, *time pressure:* Increasing time pressure by limiting the time per trial to 3 s, *rotating:* Rotating the walls clockwise after each trial) and one counterbalanced within factor (congruent and incongruent phase). Participants could only partake in the study once, as habituation effects have been shown to influence HPA axis reactivity (Roos et al., 2019; Zimmer et al., 2019). Each participant underwent both phases (congruent and incongruent), since only the incongruent phase leads to cognitive interference. The study protocol was approved by the Ethics Committee of the Friedrich-Alexander-Universität Erlangen-Nürnberg (ethical approval code 22-115-S) and adhered to the principles outlined in the Declaration of Helsinki. Additionally, the study was pre-registered on the Open Science Framework (https://osf.io/ytz7w) prior to data being collected.

2.1. Participants and design

To recruit participants, university mailing lists as well as notice boards, and social media platforms were used to reach as broad a population as possible. The following exclusion criteria were applied: 1) age below 18 or above 45 years, 2) color vision deficiency (e.g. red-green deficiency), 3) diagnosis of acute and/or chronic somatic disease, 4) use of prescription medications (especially beta blockers or glucocorticoid drugs), 5) smoker (>5 cigarettes per day; Zimmer et al., 2019), 6) prior experience with the stress protocol, 7) for female participants: hormonal contraceptives, pregnancy or menopause, 8) individuals with a body mass index <18 or >30 kg/m^2, 9) psychotherapeutic treatment in the last year, 10) regular night shift work (Niu et al., 2011). Participants were asked to refrain from eating or drinking 2 h before testing and were instructed not to drink alcohol, nor smoke or be physically active 24 h before partaking in the study. Female participants were scheduled to participate in the study during the second half of their menstrual cycle as studies show that salivary cortisol responses to a stressor in the luteal phase are comparable to the responses found in men (Kirschbaum et al., 1999). All participants gave written informed consent prior to participation.

The final sample consisted of 89 healthy men and women between 18 and 45 years (52 women; age in years: 23.60 ± 3.88), assigned randomly and stratified by sex to three configurations of the VRSR. The distribution across the conditions was roughly equal, with 31 participants (18 women) in the regular condition and 29 (17 women) in each the condition with additional time pressure and the one with additional spatial difficulty.

To assess participants' stress experience over the past month, the German version of the Perceived Stress Scale (PSS; Klein et al., 2016) was used. This questionnaire consists of 10 items that are answered on a 5-point Likert scale with answers ranging from 0 ("never") to 4 ("very often"). With a Cronbach's α of 0.84 the PSS showed good internal consistency. Participants in our study showed slightly lower values in Helplessness (M = 6.67; SD = 2.72) and Self Efficacy (M = 4.40, SD = 1.89) compared to the normative sample.

The sample for sAA analyses consisted of 85 participants, as two participants did not produce enough saliva for analyses and two further samples did not result in a valid measure for either the first or last time point, thus not allowing for linear interpolation. The same was true for cortisol, with the same two participants' samples being empty and three additional ones with invalid results for the first or last sample. Further, two participants had to be excluded for elevated baseline values for cortisol (≥12 nmol/L at baseline) compared to the overall sample, thus leading to a final sample for cortisol of 82 (47 women). The sample for ECG analyses consisted of 87 participants, as data of two participants could not be evaluated due to technical difficulties during the recording.

2.2. Apparatus

The VRSR is a hexagonal room with each wall being depicted in a different color, according to the CIELAB color characterization system, while the floor and ceiling are a dark grey (Fig. 1). These configurations are chosen to ensure maximum contrast and clarity (Braun, Fairchild, & Ebner, 1998). Participants are placed in the middle of the room via a head-mounted display (HMD) and are shown instructions which change trial by trial. Since the VRSR is conducted non-verbally, the original task of only indicating the color the word is written in was not feasible. Participants could just turn their heads until the color the word is written in and the wall color matched (MacLeod, 1991). Thus, the instructions per trial are comprised of two lines. The first line either reads WORD or COLOR and the second line consists of a color word, corresponding to the colors of the walls. The first line, the task instruction, switches in an arbitrary manner, although it is ensured that each subsequent trial does not contain either the color or word of the previous trial to monitor repetition- and successive-trial-effects (MacLeod, 1991). The baseline (*regular*) configurations allocated participants 5 s per trial, whereas the *time-pressure* condition limited their time to only 3 s. In the *rotation* condition, the walls of the room were rotated clockwise by one wall after each trial, requiring participants to constantly update their position within the room.

Fig. 1. The VRSR from the participants' point of view. VRSR = Virtual Reality Stroop Room. The time-bar depicts the time left for the participant to complete the trial. The instruction for this trial is to react to the color dimension of the color-word. In the bottom left the controller that is held by the participant can be seen with the blue line pointing at the location that is currently selected. Once the participant presses the button on the controller, this color-wall will be selected and logged and the next trial will start. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

To generate the VR environment the Steam Source engine (Valve Corporation, Bellevue, Washington, USA) was used. The VRSR was implemented in Unity 3D (Unity Technologies, San Francisco, USA). A HMD (HTC Vive Pro Eye; HTC Corporation, Taoyuan, Taiwan) and one of the HTC Vive controllers were used, allowing participants to point the virtual laser pointer and select the according wall by pressing the trigger button with the index finger of their dominant hand during the experiment. The participants' ECG was monitored and recorded with the Nexus Kit 10 MKI (Mind Media, Herten, Netherlands) and the corresponding BioTrace+ 2018 software with a sampling rate of 256 Hz.

2.3. *Measures*

2.3.1. *Task completion times and errors*

Task completion time (in seconds) in the VRSR was measured as the time between stimulus presentation and making a selection with the controller or timeout. An error was either the selection of a wall in a wrong color, failing to make a selection within the given time limit, or selecting the floor or ceiling.

2.3.2. *Questionnaires*

In addition to the standard demographic and health-related data such as sex, age, education, body mass index, previous illnesses and medication, data from psychological questionnaires was collected.

To assess positive and negative affect changes, the Positive and Negative Affect Schedule (PANAS; Krohne et al., 1996) was administered before and after participation in the experiment. It consists of 20 items, with 10 items assessing positive affect and 10 items measuring negative affect. Items are answered on a 5-point Likert Scale ranging from 1 ("very inapplicable") to 5 ("very applicable"). Both scales, positive and negative affect, show good internal consistency with a Cronbach's α of 0.84.

To our knowledge no German version of the Short Stress State Questionnaire (SSSQ; Helton & Näswall, 2015) exists. Therefore, the scale was forward translated from English to German and then back translated by three bilingual speakers. The SSSQ consists of a 24-item state pre-questionnaire and a 24-item state post-questionnaire, with both versions being rated on a 5-point Likert scale with answers between 1 ("Not at all") to 5 ("Extremely"). In the English SSSQ the three factors *Engagement, Distress* and *Worry* show good internal consistency with Cronbach's α between 0.80 and 0.89.

2.3.3. *Electrocardiogram*

Participants' ECG was recorded during the experiment at a sampling frequency of 256 Hz. The raw ECG signal was filtered with a second-order FIR bandpass filter (3–45 Hz) to reduce noise and RR intervals were computed using the QRS detection algorithm proposed by Hamilton (Hamilton, 2002). RR intervals corresponding to a heart rate of <45 bpm or >200 bpm, outliers ($\geq 2.576\sigma$) and differences of successive RR intervals ($\geq 1.96\sigma$) were subsequently removed and imputed by linear interpolation. Finally, data was split into sample periods in accordance with the two phases of the experiment and mean HR (normalized to baseline) was computed.

2.3.4. *Salivary alpha-amylase and cortisol*

The wet biomarkers of interest in this study were sAA and cortisol. As blood sugar levels influence the stress-induced cortisol response, the recommendation is to avoid large variations in blood sugar levels between participants (Zänkert et al., 2020). Therefore, 200 ml of grape juice were administered to participants upon arrival at the laboratory. Samples were collected with Salivettes (Sarstedt, Nümbrecht, Germany) at five time points and were subsequently stored for later analyses at −18 °C. They were centrifuged at 2000g and 20 °C for 10 min, before the cortisol concentrations in saliva (nmol/L) were determined in the laboratory of the Chair of Health Psychology at Friedrich-Alexander-Universität Erlangen-Nürnberg, using a commercially available chemiluminescence immunoassay (CLIA, IBL-Hamburg, Hamburg, Germany). Concentrations for sAA (U/ml) were determined using an in-house assay as described previously (Hauck et al., 2022). Intra- and Inter assay CVs were below 10%.

2.4. *Procedure*

After completing the online screening to determine eligibility for the study, participants were invited to a laboratory appointment at 13:00, 15:00 or 17:00. These times were chosen to minimize the potential impact of circadian variation of cortisol (Kudielka & Wüst, 2010). Participants gave written and informed consent and were subsequently asked to answer a range of pre-experimental questionnaires and give a baseline saliva sample (−30 min), before being connected to the Nexus Kit 10 MKI, for the recording of biosignals. To ensure familiarization with the experience of being in a VR, participants were then placed in the Steam VR Home Room, looking at a mountain landscape from a terrace, for 10 min. During that time a resting ECG in a sitting position was recorded and after 5 min the second saliva sample (S1) was obtained. After this time, participants were asked to stand up and were then placed in the VRSR.

Each participant performed both phases – congruent and incongruent – of the VRSR. To ensure that order effects were controlled for, half the participants performed the congruent phase first, while the other half started with the incongruent phase. Before each phase with 120 trials, participants were presented with 10 trials to familiarize themselves with the task. Each trial lasted a maximum of 5 s with a red bar depicting the remaining time. Participants were asked to react as fast as they could while trying to minimize errors. If they did not react within the given timeframe, an error sound was played and the screen flashed red. The same was true, if they made a mistake. Between the two phases there was a 5-min break, during which participants remained in the VRSR and were shown the countdown until the start of the next phase. A third saliva sample was collected during this break (S2). After finishing the second phase of the experiment, participants were asked to provide a fourth saliva sample (S3) before the experimenter assisted them in removing the VR equipment. They then answered the post-questionnaires of the setup and gave the last saliva sample 15 min (S4) after the end of the task. Fig. 2 gives an overview over the timeline of the experiment.

2.5. *Statistical analyses*

Since neither the sAA nor cortisol data exhibited a normal distribution, we applied log-transformation to the cortisol data and square-root-transformation to the sAA data. All results and figures are based on these transformed data. Regarding the questionnaire data, three outliers were identified ($>3\sigma$ in the z-transformed data), but as these were within the possible range of the respective questionnaire scales, we decided to include them in all analyses and to perform the analyses excluding the respective outliers for control purposes only.

We conducted mixed ANOVAs on physiological, behavioural and subjective data, testing for the effects of time as the within-subject factor, and condition (regular, time pressure, rotation), order of the phases (congruent phase first, incongruent phase first) and sex (male, female) as the between-subject factors. Data analysis was conducted in the order of data recording. Given that participants either underwent the study with the congruent or the incongruent phase first, Phase 1 and Phase 2 consist of data from both congruent and incongruent trials, contingent upon the sequence in which participants completed these phases. For HR, task completion times and errors, data were grouped by phase (congruent, incongruent) to facilitate interpretability of the results.

If Mauchly's test indicated violation of the assumption of sphericity, Greenhouse-Geisser correction was used. All analyses were conducted using Jupyter Lab (Version 3.4.4) in Python (Version 3.9.13), utilizing the open-source packages pingouin (v0.5.3; Vallat, 2018 and BioPsyKit

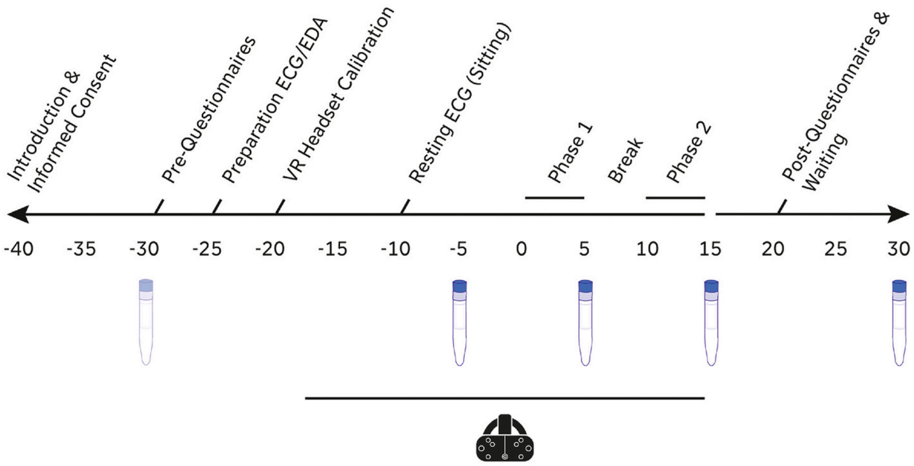

Fig. 2. Timeline of the VRSR Experiment. VRSR = Virtual Reality Stroop Room. ECG = Electrocardiogram. EDA = Electrodermal activity. Phase 1 and Phase 2 both consist of congruent or incongruent trials, depending on the order in which the participants underwent the experiment.

(v0.10.2; Richer, Küderle, Ullrich, Rohleder, & Eskofier, 2021), as well as using IBM SPSS Statistics 29 (Chicago, Illinois, USA). Significance levels were set at $p < .05$ and pairwise comparisons were Bonferroni corrected. Effect sizes are reported as partial η^2.

3. Results

3.1. Task completion time and errors

The mixed ANOVA for task completion time (TCT; seconds) revealed a significant effect of time ($F(1,77) = 1000.73$, $p < .001$, $\eta p^2 = 0.93$; congruent phase: $M = 1.51$, $SD = 0.32$; incongruent phase: $M = 2.54$, $SD = 0.38$) and a significant interaction for time × condition ($F(2,77) = 8.53$, $p < .001$, $\eta p^2 = 0.18$) and time × order ($F(1,77) = 13.34$, $p < .001$, $\eta p^2 = 0.15$). Participants exhibited shorter TCT under the *time-pressure* condition (*regular*: $M = 2.15$, $SD = 0.68$; *time pressure*: $M = 1.78$, $SD = 0.46$; *rotating*: $M = 2.13$, $SD = 0.64$). Those who experienced the congruent phase first demonstrated longer TCT within that phase ($M = 1.58$, $SD = 0.34$) but shorter TCT within the incongruent phase ($M = 2.50$, $SD = 0.38$), compared to those who experienced the experiment in the reverse order (congruent: $M = 1.44$, $SD = 0.27$; incongruent: $M = 2.58$, $SD = 0.37$). All other effects were n.s. (smallest $p = .09$).

With regard to errors, the mixed ANOVA showed a main effect of time ($F(1,77) = 102.16$, $p < .001$, $\eta p^2 = 0.57$), indicating that participants made more errors in the incongruent phase, as well as an interaction effect for time × order × sex ($F(1,77) = 5.39$, $p = .02$, $\eta p^2 = 0.07$), indicating men made more mistakes in the incongruent condition compared to women when they underwent the experiment with the congruent phase first, whereas the opposite was true for the group of participants starting with the incongruent condition (congruent phase first: women congruent: $M = 1.78$, $SD = 2.42$; women incongruent: $M = 9.52$, $SD = 7.07$; men congruent: $M = 1.21$, $SD = 2.78$; men incongruent: $M = 15.11$, $SD = 15.92$; incongruent phase first: women congruent: $M = 3.56$, $SD = 9.21$; women incongruent: $M = 16.16$, $SD = 12.20$; men congruent: $M = 1.11$, $SD = 1.75$; men incongruent: $M = 10.11$, $SD = 10.59$). All other effects were n.s. (smallest $p = .30$).

3.2. Questionnaires

3.2.1. Short Stress State Questionnaire

For the pre-experimental measurement, the factors *Engagement* ($M = 3.59$, $SD = 0.46$) and *Worry* ($M = 2.32$, $SD = 0.78$) corresponded with the values of the original paper (Helton & Näswall, 2015), while *Distress* showed lower values ($M = 1.27$, $SD = 0.35$). The same was true for the post-experimental questionnaire data for *Engagement* ($M = 3.83$, $SD = 0.58$), Worry ($M = 2.05$, $SD = 0.73$) and *Distress* ($M = 1.42$, $SD = 0.43$). The mixed ANOVA revealed a significant effect of time for all three scales (*Engagement:* $F(1,77) = 29.46$; $p < .001$; $\eta p^2 = 0.28$; *Worry:* $F(1,77) = 20.60$; $p < .001$; $\eta p^2 = 0.21$; *Distress:* $F(1,77) = 8.36$; $p = .005$; $\eta p^2 = 0.10$), indicating that the VRSR induced a rise in *Engagement* and *Distress*, while *Worry* was lower after the experiment. For *Engagement* we also found a time × sex × condition interaction ($F(2,77) = 9.81$; $p < .001$; $\eta p^2 = 0.20$), men and women differed with regard to their stress-induced change of engagement ratings in the three conditions. All other effects were n.s. (smallest $p = .18$) and excluding the three outliers did not result in significant changes regarding the results.

3.2.2. Positive affect Negative Affect Schedule

The PANAS was used as a pre and post measure. Positive Affect (PA) was higher after the experiment (pre: $M = 2.91$, $SD = 0.61$; post: $M = 3.05$, $SD = 0.77$), while Negative Affect (NA) showed lower values post-experiment (pre: $M = 1.32$, $SD = 0.39$; post: $M = 1.24$, $SD = 0.27$). The mixed ANOVA revealed a main effect of time for PA ($F(1,77) = 5.00$; $p = .03$; $\eta p^2 = 0.06$) and NA ($F(1,77) = 7.22$; $p = .01$; $\eta p^2 = 0.09$). For NA, the data also showed a significant time × sex interaction ($F(1,77) = 5.41$; $p = .02$; $\eta p^2 = 0.07$). However, this interaction was no longer significant when excluding outliers. All other effects were n.s. (smallest $p = .32$) and excluding the three outliers did not result in further changes regarding the results. These results indicate that participants positive affect was higher after the experiment, while their negative affect decreased.

3.3. Electrocardiogram

For HR (bpm) the mixed ANOVA revealed a significant effect of time ($F(1, 75) = 15.70$, $p < .001$, $\eta p^2 = 0.17$) indicating that participants showed a rise in HR in the incongruent phase. Further, a significant time × condition interaction ($F(2, 75) = 3.60$, $p = .03$, $\eta p^2 = 0.09$) emerged, indicating that participants HR in the congruent and incongruent phase differed across the three conditions (Fig. 3). All other effects were n.s. (smallest $p = .06$).

3.4. Salivary alpha-amylase

For sAA, we conducted a mixed ANOVA which revealed a significant effect of time ($F(2.40, 175.49) = 4.92$, $p = .01$, $\eta p^2 = 0.06$). All other effects were n.s. (smallest $p = .22$), indicating that participants showed an increase in sAA over the course of the VRSR, while neither the order of the phases, nor the conditions did influence sAA concentrations (Fig. 4).

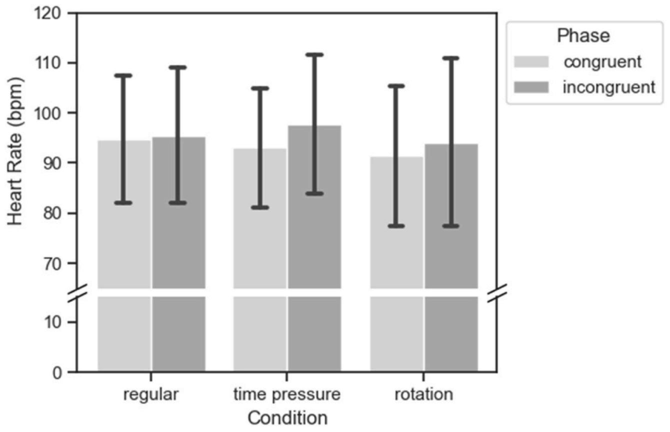

Fig. 3. Heart rate for the conditions of the VRSR divided by the congruent and incongruent phase. Error bars show standard deviation (SD).

3.5. Cortisol

For cortisol, the mixed ANOVA revealed a significant time × sex × order interaction ($F(2.05,143.36) = 5.44$, $p = .01$, $\eta p^2 = 0.07$), as well as a significant time × sex × order × condition interaction ($F(4.10,143.36) = 3.07$, $p = .02$, $\eta p^2 = 0.08$). These results indicate that men and women differed in their cortisol response across time, not only with regard to the order of congruent and incongruent phase, but also in their reaction to the three conditions (Fig. 5). All other effects were n.s. (smallest $p = .20$).

4. Discussion

This study shows that the VRSR, despite its inherent differences to the original Stroop Color and Word Test (SCWT), results in prolonged task completion times (TCT) and increased error rates in the incongruent phase compared to the congruent phase. Thus, the VRSR can be used to replicate these specific effects from the SCWT, while it also allows the assessment of EF as the SCWT is a measure for participants' inhibition ability, while the three-dimensionality and the added task-switching

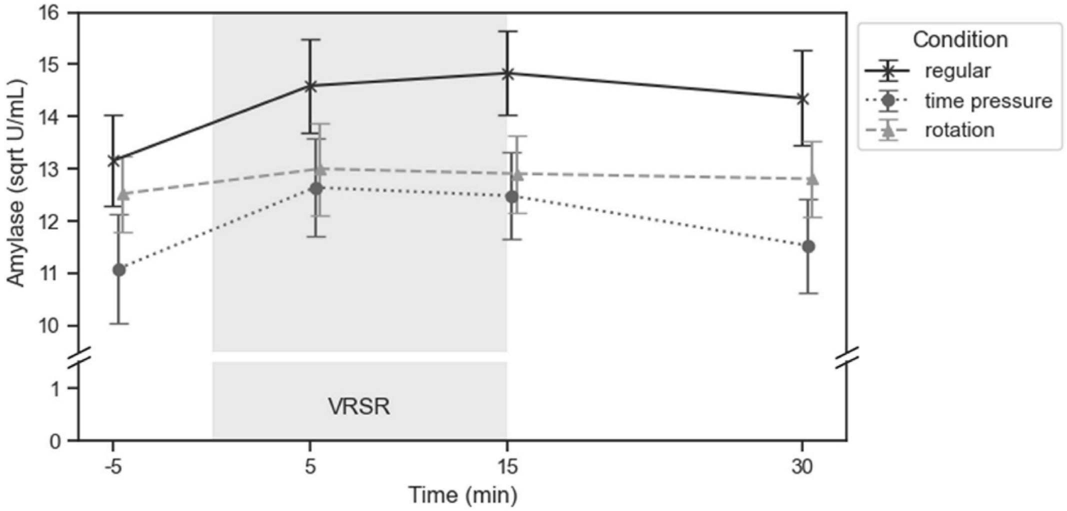

Fig. 4. Salivary alpha-amylase for the conditions of the VRSR over the four collection time points before and after the VRSR. Figure shows square root transformed values of sAA, with error bars depicting standard deviations (SD).

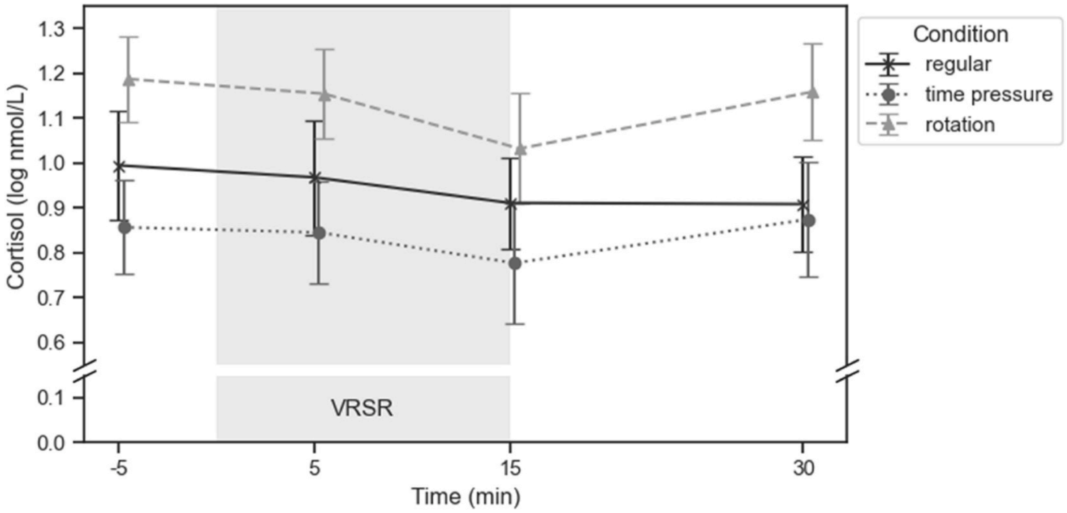

Fig. 5. Salivary cortisol for the conditions of the VRSR over the four collection time points before and after the VRSR. Figure shows log-transformed values of salivary cortisol, with error bars depicting standard deviations (SD).

component of the VRSR also tap into updating and cognitive flexibility respectively. With VR becoming more and more prevalent, this opens many possibilities with regard to research, training and intervention. VR environments can be adapted with low cost and effort, and provide a high degree of control over nearly all aspects of participants' environment.

A more detailed look into TCT and errors shows that participants who only had 3 s per trial (*time-pressure* condition) exhibited shorter TCT, but not more errors. As participants do not report higher subjective stress in the SSSQ, nor do they make more errors, we conclude that time in the VRSR can be adjusted to 3 s, which should also heighten the overall pressure on participants. Another important finding is that participants exhibited longer TCT in the phase that they underwent first, compared to the other group. A possible explanation is that participants show learning effects, such as that they need to adjust to the overall environment of the VRSR and the best strategy of handling the task at hand. Future studies should examine these effects more closely.

With regard to errors, our data points towards a difference between men and women when it comes to the two phases – congruent and incongruent. Women appear to make fewer mistakes than men in the incongruent condition when they start with the easier, congruent condition, while they make more mistakes, compared to men, when they start with the more difficult, incongruent condition.

Another query our study aimed to answer was, whether the VRSR leads to a physiological or psychological stress response. For the self-report measures of stress and affect we found ambiguous effects. Participants reported higher *Engagement* and *Distress*, as well as lower *Worry* in the post-questionnaires. Yet, they also exhibited higher PA and lower NA after the experiment. It is possible that the VRSR is stressful, but at the same time has enjoyable, game-like characteristics, leading to what might be termed *eustress* (Bienertova-Vasku et al., 2020; Kupriyanov & Zhdanov, 2014). Other theories have classified this form of stress as a challenge, as opposed to harm or threat (Lazarus & Folkman, 1984). The fact that *Engagement* and PA rise after participating in the VRSR show that it is an enjoyable experiment, in contrast to experiments that elicit physical harm or social-evaluative threat (Kirschbaum et al., 1993; Schwabe et al., 2008; Zimmer et al., 2019). Also, these findings tie in with previous results that moderate stress might facilitate engagement with a task (Shields et al., 2016) and that this in turn might be optimal for good performance as indicated by the Yerkes-Dodson-Law (Welford, 1973; Yerkes & Dodson, 1908).

The results for the physiological stress response are slightly more complex. For SNS reactivity, we assessed HR and sAA. For HR we found that the VRSR elicits a significant response, with a difference between the two experimental phases, namely congruent and incongruent. From this, we conclude that the incongruent phase of the VRSR elicits a physiological stress response via the SNS aside from just movement effects and the general task of searching for and selecting of the corresponding color wall. This is consistent with previous findings, such as the observation that higher cognitive workload leads to an increase in response level (Parsons & Courtney, 2018). What is more, we found a significant interaction effect for HR across the three conditions, revealing that for the *regular* condition the difference in HR between the congruent and incongruent phase was virtually non-existent, whereas for the *time-pressure* and *rotation* condition, the HR was higher in the incongruent phase, compared to the congruent phase. Unfortunately, for sAA, we did not find a difference between the experimental phases. However, with HR showing faster reactivity than sAA, it is possible that our set-up, with 120 trials per phase, was not optimal to find an effect for sAA as this marker reacts more slowly. Another explanation might be that HR is a better marker for faster responding facets of the SNS than sAA. In a recent meta-analysis, Dammen et al. (2022) report medium effect sizes for HR over 42 studies and only a small effect size for sAA, although only four studies were taken into account for this marker. The results for cortisol indicate that in this sample there was an interaction effect with regard to time, sex, order of the two phases, and the three conditions. However, our findings do not support a uniform response of the HPA axis to this specific laboratory protocol. Also, we are fully aware of the fact that such an interaction needs to be interpreted very carefully, given our sample size.

With this study, we aimed to look into the effects of the VRSR on measures of EF and stress. While a first evaluation with a smaller sample (Gradl et al., 2019) resulted in a significant cortisol effect for some of the conditions, in our larger sample we were unable to replicate this result. However, with regard to HR our findings support that this fast-reacting marker for the SNS differs across the congruent and incongruent phase of the VRSR. Salivary alpha-amylase – another indicator for the SNS – did not show the same pattern, which we attribute to the study design and set-up. We were further able to reproduce the effects of the SCWT with regard to TCT and errors, thereby demonstrating that the VRSR is comparable to the SCWT in these basic effects. Questionnaire data supports our claim that the VRSR is a mild stressor, yet with an enjoyable, game-like quality, which does not result in a rise of negative affect, but on the contrary in a rise of positive affect and engagement with the task.

In evaluating the influence of the three conditions (*regular, time pressure,* and *rotation*), our analyses generally indicate only a few substantial differences. Participants exhibited notably quicker TCT under the *time-pressure* condition, without a corresponding increase in errors compared to other conditions. This observation suggests the feasibility of implementing a 3-s time limit for the entire test, as participants have demonstrated their ability to perform the task within this shorter timeframe. Further, for the HR we also found a condition effect, suggesting that more time pressure or more spatial difficulty influence this marker of the SNS.

Compared to the original study (Gradl et al., 2019), it has to be noted that we introduced some changes to the set-up. First, we changed the colors from the default Unity colors to the CIELAB color system. This was done to achieve maximum contrast and clarity (Braun et al., 1998). Second, in the original implementation the walls in condition C, corresponding to the *rotation* condition in our version, did not rotate one wall after each trial, but would be swapped with the opposite wall after each trial. Also, the color field would get smaller with each trial so that even having located the right wall, participants still faced the challenge of selecting the now smaller target area with the controller. If they selected the right wall without hitting the target area, this would count as a mistake. Our goal in rearranging the room after each trial was not to measure or challenge motor skills, but to engage the participants' ability to mentally adapt to a changing environment and update their location relative to the color walls. Third, we made sure that repetition- and successive-trial-effects (MacLeod, 1991) were controlled for by ensuring that neither the word nor the color from one trial would be part of the next trial. Finally, we anticipated an increase in statistical power, as our sample was considerably larger with 89 participants and three conditions, as compared to 32 participants and four conditions (Gradl et al., 2019).

While this study demonstrates several strengths, it is important to acknowledge the following limitations. First, the results regarding the SNS might have been confounded by the movement of participants. While they do not physically exert themselves, they still rotate their upper body, in some cases even their whole body, to find the correct wall. However, as HR was higher in the incongruent phase, regardless of the phase participants underwent first, we can still draw the conclusion that the incongruent phase of the VRSR, activated the SNS aside from these movement artifacts. Second, the VRSR does, in its current set-up, not allow for assessment of reaction times, defined as the time from presenting a stimulus to the initiation of a reaction of the participant. Instead, we measured task completion time (TCT), which is the time from presentation of the stimulus to selecting a wall. This measure includes the time spent searching for the correct wall. In addition, the incongruent phase requires choice as well as inhibition processes to be activated and both play a role with regard to the time it takes from stimulus presentation to starting to search for the corresponding wall (Logan & Cowan, 1984; Stone, 1960). Participants do not only have to choose which dimension of the color word they pay attention to they also have to inhibit the other feature. The current version of VRSR cannot distinguish between these processes. Nevertheless, our results with regard to TCT (and errors) are in accordance with the Stroop Effect. Participants take longer (and make more mistakes) in the incongruent phase. Further limitations include the young sample, mostly consisting of university students, and the fact that the VRSR does not allow the assessment of a singular EF facet, such as updating or shifting, but captures an overall EF factor (Miyake et al., 2000). As cognitive functions diminish with age (Glisky, 2007), future studies would benefit from trying to reach a wider sample with regard to age. Also, we would be interested to see how other researchers might adapt the VRSR with regard to specific EF components or potentially stressful components.

In summary, our findings endorse the VRSR as an enjoyable yet challenging task that elicits a stress response via the SNS, yet not the HPA axis, and shows a similar pattern as the SCWT with regards to errors and processing times. Self-report data underscore these findings. Hence, our study successfully fulfills its objective of assessing the VRSR utilizing a large sample, offering a foundation for the design and execution of future investigations.

Pre-registration

Ringgold, V., & Rohleder, N. (2022, August 24). The Virtual Reality Stroop Room: Cognitive Stress in Virtual Reality. Retrieved from https://doi.org/10.17605/OSF.IO/YTZ7W

Publication ethics

The study was conducted in accordance with the Declaration of Helsinki. All participants gave written informed consent before testing. The Ethics Committee of the medical faculty of the Friedrich-Alexander-Universität Erlangen-Nürnberg gave ethical approval for the study that was conducted (ethical approval code 22-115-S).

Funding

This work was (partly) funded by the Konrad-Adenauer-Stiftung and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1483 – Project-ID 442419336, EmpkinS.

CRediT authorship contribution statement

Veronika Ringgold: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Luca Abel:** Data curation, Visualization. **Bjoern M. Eskofier:** Funding acquisition, Resources, Supervision, Writing – review & editing. **Nicolas Rohleder:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare no conflicts of interest.

Data availability

Data will be made available on request.

List of Abbreviations

ANS	Autonomous Nervous System
ECG	Electrocardiogram
EF	Executive Functions
HMD	Head-Mounted-Display
HPA	Hypothalamic-Pituitary-Adrenal
HR	Heart Rate
PANAS	Positive and Negative Affect Schedule
PNS	Parasympathetic Nervous System
PSS	Perceived Stress Scale
sAA	Salivary alpha-amylase
SCWT	Stroop Color and Word Test
SNS	Sympathetic Nervous System
SSSQ	Short Stress State Questionnaire
TSST	Trier Social Stress Test
VR	Virtual Reality
VRSR	Virtual Reality Stroop Room

References

Bienertova-Vasku, J., Lenart, P., & Scheringer, M. (2020). Eustress and distress: Neither good nor bad, but rather the same? *BioEssays, 42*(7), Article 1900238. https://doi.org/10.1002/bies.201900238

Braun, G. J., Fairchild, M. D., & Ebner, F. (1998). Color gamut mapping in a hue-linearized CIELAB color space. *Color and Imaging Conference, 6*, 163–168.

Cohen, S., Janicki-Deverts, D., & Miller, G. E. (2007). Psychological stress and disease. *JAMA, 298*(14), 1685. https://doi.org/10.1001/jama.298.14.1685

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

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Fauvel, J. P., Bernard, N., Laville, M., Daoud, S., Pozet, N., & Zech, P. (1996). Reproducibility of the cardiovascular reactivity to a computerized version of the Stroop stress test in normotensive and hypertensive subjects. *Clinical Autonomic Research, 6*(4), 219–224. https://doi.org/10.1007/BF02291137

Glisky, E. L. (2007). Changes in cognitive function in human aging. In *Brain aging: Models, methods, and mechanisms*. CRC Press/Taylor & Francis.

Gradl, S., Wirth, M., Mächtlinger, N., Poguntke, R., Wonner, A., Rohleder, N., & Eskofier, B. M. (2019). The Stroop room: A virtual reality-enhanced Stroop test. *25th ACM symposium on virtual reality software and technology*. https://doi.org/10.1145/3359996.3364247

Grant, K. E., Compas, B. E., Thurm, A. E., McMahon, S. D., & Gipson, P. Y. (2004). Stressors and child and adolescent psychopathology: Measurement issues and prospective effects. *Journal of Clinical Child and Adolescent Psychology, 33*(2), 412–425.

Hamilton, P. (2002). Open source ECG analysis. *Computers in Cardiology*, 101–104. https://doi.org/10.1109/CIC.2002.1166717

Hassard, J., Teoh, K. R. H., Visockaite, G., Dewe, P., & Cox, T. (2018). The cost of work-related stress to society: A systematic review. *Journal of Occupational Health Psychology, 23*(1), 1–17. https://doi.org/10.1037/ocp0000069

Hauck, F., Romero Gibu, L., Jansen, S., & Rohleder, N. (2022). Differences in acute stress responses depending on first or second language in a Hispanic-American sample. *Stress: The International Journal on the Biology of Stress, 25*(1), 313–322. https://doi.org/10.1080/10253890.2022.2110466

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology, 34*(2), 163–171. https://doi.org/10.1016/j.psyneuen.2008.10.026

Helton, W. S., & Näswall, K. (2015). Short stress state questionnaire. *European Journal of Psychological Assessment, 31*(1), 20–30.

Henderson, R. K., Snyder, H. R., Gupta, T., & Banich, M. T. (2012). When does stress help or harm? The effects of stress controllability and subjective stress response on Stroop performance. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00179

Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research, 11*(2), 213–218. https://doi.org/10.1016/0022-3999(67)90010-4

Kagan, J. (2016). An overly permissive extension. *Perspectives on Psychological Science, 11*(4), 442–450. https://doi.org/10.1177/1745691616635593

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin, 144*(11), 1147.

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. *Psychosomatic Medicine, 61*(2), 154–162. https://doi.org/10.1097/00006842-199903000-00006

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

Klein, E. M., Brähler, E., Dreier, M., Reinecke, L., Müller, K. W., Schmutzer, G., Wölfling, K., & Beutel, M. E. (2016). The German version of the Perceived Stress Scale – psychometric characteristics in a representative German community sample. *BMC Psychiatry, 16*(1), 159. https://doi.org/10.1186/s12888-016-0875-9

Krohne, H., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der “Positive and Negative Affect Schedule” (PANAS). *Diagnostica, 42*, 139–156. https://doi.org/10.1037/t49650-000

Kudielka, B. M., & Wüst, S. (2010). Human models in acute and chronic stress: Assessing determinants of individual hypothalamus–pituitary–adrenal axis activity and reactivity. *Stress: The International Journal on the Biology of Stress, 13*(1), 1–14. https://doi.org/10.3109/10253890902874913

Kupriyanov, R., & Zhdanov, R. (2014). The eustress concept: Problems and outlooks. *World Journal of Medical Sciences, 11*(2), 178–185.

Kuras, Y. I., McInnis, C. M., Thoma, M. V., Chen, X., Hanlin, L., Gianferante, D., & Rohleder, N. (2017). Increased alpha-amylase response to an acute psychosocial stress challenge in healthy adults with childhood adversity. *Developmental Psychobiology, 59*(1), 91–98. https://doi.org/10.1002/dev.21470

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping.* Springer.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit simple and choice reaction time responses: A model and a method. *Journal of Experimental Psychology, 10*(2), 276–291.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203.

McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences, 840*(1), 33–44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x

McEwen, B. S., & McEwen, C. A. (2016). Response to jerome kagan’s essay on stress (2016). *Perspectives on Psychological Science, 11*(4), 451–455. https://doi.org/10.1177/1745691616646635

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin, 133*(1), 25–45. https://doi.org/10.1037/0033-2909.133.1.25

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Niu, S.-F., Chung, M.-H., Chen, C.-H., Hegney, D., O’Brien, A., & Chou, K.-R. (2011). The effect of shift rotation on employee cortisol profile, sleep quality, fatigue, and attention level: A systematic review. *Journal of Nursing Research, 19*(1), 68–81. https://doi.org/10.1097/JNR.0b013e31820c1879

Parsons, T. D., & Courtney, C. G. (2018). Interactions between threat and executive control in a virtual reality Stroop task. *IEEE Transactions on Affective Computing, 9*(1), 66–75. https://doi.org/10.1109/TAFFC.2016.2569086

Quinn, M. E., & Shields, G. S. (2023). The insidious influence of stress: An integrated model of stress. *Executive Control, and Psychopathology. Clinical Psychological Science, 11*(5), 773–800. https://doi.org/10.1177/21677026221149736.

Renaud, P., & Blondin, J.-P. (1997). The stress of Stroop performance: Physiological and emotional responses to color–word interference, task pacing, and pacing speed. *International Journal of Psychophysiology, 27*(2), 87–97. https://doi.org/10.1016/S0167-8760(97)00049-4

Richer, R., Küderle, A., Ullrich, M., Rohleder, N., & Eskofier, B. (2021). BioPsyKit: A Python package for the analysis of biopsychological data. *Journal of Open Source Software, 6*(66), 3702.

Rohleder, N. (2012). Acute and chronic stress induced changes in sensitivity of peripheral inflammatory pathways to the signals of multiple stress systems – 2011 Curt Richter Award Winner. *Psychoneuroendocrinology, 37*(3), 307–316. https://doi.org/10.1016/j.psyneuen.2011.12.015

Rohleder, N. (2019). Stress and inflammation – the need to address the gap in the transition between acute and chronic stress effects. *Psychoneuroendocrinology*. https://doi.org/10.1016/j.psyneuen.2019.02.021

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

Scarpina, F., & Tagini, S. (2017). The Stroop color and word test. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.00557

Schneiderman, N., Ironson, G., & Siegel, S. D. (2005). Stress and health: Psychological, behavioral, and biological determinants. *Annual Review of Clinical Psychology, 1*(1), 607–628. https://doi.org/10.1146/annurev.clinpsy.1.102803.144141

Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology, 33*(6), 890–895. https://doi.org/10.1016/j.psyneuen.2008.03.001

Shields, G. S. (2020). Stress and cognition: A user’s guide to designing and interpreting studies. *Psychoneuroendocrinology, 112*, Article 104475. https://doi.org/10.1016/j.psyneuen.2019.104475

Shields, G. S., Bonner, J. C., & Moons, W. G. (2015). Does cortisol influence core executive functions? A meta-analysis of acute cortisol administration effects on working memory, inhibition, and set-shifting. *Psychoneuroendocrinology, 58*, 91–103. https://doi.org/10.1016/j.psyneuen.2015.04.017

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews, 68*, 651–668. https://doi.org/10.1016/j.neubiorev.2016.06.038

Shields, G. S., & Slavich, G. M. (2017). Lifetime stress exposure and health: A review of contemporary assessment methods and biological mechanisms. *Social and Personality Psychology Compass, 11*(8), Article e12335. https://doi.org/10.1111/spc3.12335

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00328

Stone, M. (1960). Models for choice-reaction time. *Psychometrika, 25*(3), 251–260. https://doi.org/10.1007/BF02289729

Strahler, J., Skoluda, N., Kappert, M. B., & Nater, U. M. (2017). Simultaneous measurement of salivary cortisol and alpha-amylase: Application and recommendations. *Neuroscience & Biobehavioral Reviews, 83*, 657–677. https://doi.org/10.1016/j.neubiorev.2017.08.015

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643.

Tulen, J. H. M., Moleman, P., van Steenis, H. G., & Boomsma, F. (1989). Characterization of stress reactions to the Stroop color word test. *Pharmacology Biochemistry and Behavior, 32*(1), 9–15. https://doi.org/10.1016/0091-3057(89)90204-9

Vallat, R. (2018). Pingouin: Statistics in python. *Journal of Open Source Software, 3*, 1026. https://doi.org/10.21105/joss.01026.

Welford, A. T. (1973). Stress and performance. *Ergonomics, 16*(5), 567–580. https://doi.org/10.1080/00140137308924547

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Zänkert, S., Kudielka, B. M., & Wüst, S. (2020). Effect of sugar administration on cortisol responses to acute psychosocial stress. *Psychoneuroendocrinology, 115*, Article 104607. https://doi.org/10.1016/j.psyneuen.2020.104607

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