DISSERTATION

A Novel Approach to Inducing and Measuring Acute Stress Responses: Development, Evaluation and Application Potential of a Digital Stress Test

Ein innovativer Ansatz für die Induktion und Messung von Akuten Stressreaktionen: Entwicklung, Evaluierung und Anwendungspotential eines Digitalen Stress Tests

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät

Charité – Universitätsmedizin Berlin

von

Matthias Norden

Erstbetreuung: Prof. Dr. med. Dr. rer. nat. Felix Balzer

Datum der Promotion: 30. Juni 2024

Inhaltsverzeichnis

List of Tables

List of Figures

List of Abbreviations

ANOVA	analysis of variances
API	application programming interface
C-DST	Control - Digital Stress Test
DST	Digital Stress Test
GUI	graphical user interface
HPA	hypothalamic–pituitary–adrenal
HTTPS	Hyper Text Transfer Protocol Secure
i-PANAS SF	international short form of the Positive And Negative Affect Schedule
MIST	Montreal Imaging Stress Task
NA	negative affect
n/a	not applicable
PA	positive affect
PANAS	Positive And Negative Affect Schedule
SCP	secure copy protocol
SSL/TLS	secure sockets layer/transport layer security
VAS	visual analogue scale
VR	virtual reality

Abstract

Acute stress is linked to a variety of negative outcomes, including increased risk for mental and physical diseases, and reduced quality of life. Effective induction and accurate measurement of acute stress responses are important for both research and clinical purposes. Traditional methods rely on laboratory-based stressors, which can be costly, time-consuming, and impractical for large-scale studies or real-world applications. Measurements in outside-the-lab settings mostly reflect subjective stress levels while objective and feasible measures of biological stress consequences are scarce. This thesis aims to overcome these limitations by linking traditional psychological stress research with innovative computer science methods.

First, covered by a published study, the concept, development and online evaluation of a new Digital Stress Test (DST) for the induction and video-recording of acute stress responses are presented. In this study, the first prototype of the DST was tested in a large and experimenter-independent online study with 284 participants. Results show that the DST could induce significantly higher levels of perceived stress and negative affect compared to the control condition. Going beyond this study, further developments of the DST and a pre-registered follow-up validation study are outlined. In this study, participants perform the DST and the gold standard laboratory stress induction paradigm Trier Social Stress Test while their physiological stress responses are evaluated. Lastly, the potentials of using the DST to contribute to the development of video-based stress detection methods are critically reviewed. Therefore, a follow-up online study for collecting a video dataset is outlined and, based on the results of a further already published study, the applicability of baseline machine learning algorithms for video-based stress detection discussed.

The findings in this thesis imply several potentials of the Digital Stress Test: First, the DST is applicable as a tool for inducing acute stress responses in outside-the-lab settings and thus making more ecologically valid and scalable stress studies possible. Secondly, it also allows for gathering videos capturing stress-related behavioral data in real-world scenarios and therefore supporting the development of reliable stress detection algorithms. Finally, this thesis may present the DST as an invitation for promoting open and collaborative research in the interdisciplinary field between psychology and computer science.

Zusammenfassung

Akuter Stress ist mit einer Vielzahl negativer Auswirkungen verbunden, einschließlich eines erhöhten Risikos für psychische und körperliche Erkrankungen sowie reduzierter Lebensqualität. Eine wirksame Induktion und genaue Messung akuter Stressreaktionen ist sowohl für Forschungs- als auch für klinische Zwecke relevant. Traditionelle Methoden setzen auf im Labor durchgeführte Stressoren, die kostenintensiv, zeitaufwendig und unpraktisch für groß angelegte Studien oder Anwendungen im alltäglichen Leben sein können. Messungen außerhalb des Labors spiegeln meist das subjektive Stresslevel wider, während objektive und alltagstaugliche Methoden zur Messung von biologischen Stressfolgen fehlen. Diese Dissertation zielt darauf ab, diese Einschränkungen durch die Verbindung traditioneller psychologischer Stressforschung mit innovativen Methoden der Informatik zu überwinden. Zunächst wird die veröffentlichte Studie über das Konzept, die Entwicklung und die Online-Evaluation eines neuen Digitalen Stress Tests (DST) für die Induktion und Videoaufzeichnung akuter Stressreaktionen vorgestellt. In dieser Studie wurde der erste Prototyp des DST in einer großen und experimentatorunabhängigen Online-Studie mit 284 Teilnehmenden getestet und konnte im Vergleich zur Kontrollbedingung signifikant stärkeren wahrgenommenen Stress und negativen Affekt auslösen.

Über die Studie hinausgehend werden Weiterentwicklungen des DST und eine prä-registrierte Validierungsstudie skizziert. In dieser zusätzlichen Studie führen die Teilnehmenden den DST und das Goldstandard-Stressinduktionsparadigma Trier Social Stress Test durch, wobei Daten zu physiologischen Stressreaktionen erhoben werden. Abschließend wird das Potential, den DST für die Entwicklung von videobasierten Stresserkennungsalgorithmen zu nutzen, kritisch überprüft. Dafür werden Pläne einer weiteren Online-Studie zur Erstellung eines Videodatensatzes skizziert und, basierend auf den Ergebnissen einer weiteren bereits veröffentlichen Studie, die Anwendbarkeit von Grundlagenalgorithmen des maschinellen Lernens für die videobasierte Stresserkennung diskutiert.

Die Ergebnisse dieser Dissertation zeigen die vielfältigen Einsatzmöglichkeiten des DST auf: Zunächst kann der DST zur Induktion akuter Stressreaktionen außerhalb des Labors angewendet werden und somit ökologisch valide und skalierbare Stressstudien ermöglichen. Darüber hinaus ermöglicht er die Sammlung von Videos, die stressbezogene Verhaltensdaten in realen Szenarien erfassen und unterstützt damit die

Entwicklung von zuverlässigeren Stress-Detektionsalgorithmen. Zusammenfassend können diese Dissertation und der DST als Einladung zur Förderung offener und kollaborativer Forschung im interdisziplinären Bereich zwischen Psychologie und Informatik dienen.

1 Introduction

Psychosocial stress is a well-established risk factor for the development of cardiovascular, metabolic and mental diseases, including hypertension, diabetes, anxiety and depression (Yaribeygi et al., 2017). At the same time, stress induced morbidity and perceived stress levels have been increasing for the past two decades (Eurofound and EU-OSHA, 2014), making stress an important health and socio-economic challenge of the 21st century and highlighting the importance of further research.

Most psychological stress research paradigms so far require laboratory settings, additional resources and cannot be (easily) conducted in outside-the-lab settings which impedes more ecologically valid research. The bounding to laboratory settings has also been shown to be crucial for periods where traditional (laboratory) research is not feasible or possible at all, such as during global pandemias or wars - despite the high importance of psychological research during such periods (Pfeifer et al., 2021). Further, classical stress induction procedures mostly rely on costly resources (e.g., laboratory personnel), making them unfeasible for large scale application and hence, the collection of unbiased and representative data. Remotely feasible stress research methods have been emerging but are still laborious and require experimenter contact (Eagle et al., 2021; Gunnar, 1987; Meier et al., 2022). While these methods can be seen as a first step towards laboratory-independent studies and have been shown to be extremely helpful during global crises, they are not suited for large scale studies. A psychological paradigm that could be performed by participants independently in their own specific environment without any further equipment or setups would not only allow to transfer traditional (laboratory) stress studies to outside-the-lab settings but also to gather information from much bigger and more diverse cohorts (Kirschbaum, 2021). This would not only increase the statistical power but also open the path to inclusive and unbiased study samples.

At the same time, innovative computational analysis methods have been successfully applied for various applications with promising results for the stress research domain. These methods usually require large amounts of reliably annotated and multimodal data (Althnian et al., 2021; Wang et al., 2020). So far, stress datasets including information on non-verbal behavior (e.g., facial movements, voice, head movement) are usually small and either poorly or inconsistently annotated which leads to inconclusive results (Aigrain et al., 2018; Norden, Wolf, et al., 2022).

In the published study underlying this thesis, we have conceptualized, developed and evaluated a new digital paradigm - the Digital Stress Test (DST) - for inducing and (video-) recording acute stress responses. The DST can be conducted with smartphones via a fully automated web application and does not require any further equipment or experimenter contact. It allows for administering scalable and outside-the-laboratory stress studies while gathering non-verbal behavior data captured through video recordings.

In the following, I will emphasize the necessity of such scalable and remotely feasible stress *induction* and *recording* tool by giving a brief overview of limitations in currently used methods and presenting advances in computational stress video analysis. The Introduction section ends with pointing out the aims and hypotheses investigated in the accompanying publication. I include research questions going beyond the present publication which will be discussed on the basis of an on-going pre-registered study (Norden, Mayer, et al., 2022) and results of an additional publication (Norden, Wolf, et al., 2022). In the following sections – Methods & Results – first, the general concept, technological implementation and details of the online evaluation study are summarized before most important findings of the evaluation including unpublished results are presented. Finally, in the last section – Discussion – the results of the published study are put into context and most important implications and limitations discussed. Specifically, I will discuss the limitations in the light of follow-up analyses and current and planned developments within the DST project. The Discussion section ends with ethical considerations where the presented research is critically evaluated regarding ethical concerns.

1.1 Current Stress Paradigms and their Limitations

Most *classical* stress induction paradigms are introduced and explained with respect to their limitations in the introduction section of the accompanying publication. These include the Trier Social Stress Test (TSST; Kirschbaum et al., 1993) with its adaptations for groups or virtual reality (VR) usage (Jönsson et al., 2010; Montero-López et al., 2016; Shiban et al., 2016), the Montreal Imaging Stress Task (MIST; Dedovic et al., 2005), the Imaging Paradigm for Evaluative Social Stress (Fehlner et al., 2020) and the socially evaluated cold pressor test (Minkley et al., 2014). Many of these are limited in their scalability due to being bound to laboratory settings and equipment (e.g., TSST, VR-

TSST) or have not been tested outside the lab (e.g., MIST). Only most recently – accelerated by the COVID-19 pandemia with its implications for limited possibilities of traditional face to face research – an interest in remotely feasible stress induction paradigms has emerged (Kirschbaum, 2021; Pfeifer et al., 2021). Several versions of online (video-conferencing) versions of the TSST have been tested since its first publications by Gunnar et al. (2021) and Eagle et al. (2021). For the online TSST, participants perform the mock job interview task and mental arithmetic task in front of a videoconferencing tool without any in-person assessments. Various studies showed that online TSSTs can induce subjective stress, and activate the autonomic nervous system (Harvie et al., 2021; Huneke et al., 2021; Reed et al., 2023). Further, recent studies (Gunnar et al., 2021; Meier et al., 2022) also showed the efficacy of the online versions in activating the HPA axis as measured in saliva cortisol. In their review on stress research during the COVID-19 pandemic, Pfeifer et al. (2021) also illustrate the potentials and perspectives of remotely feasible stress induction paradigms going beyond the pandemic. First, such procedures would increase the accessibility and enable standardized research in field contexts (e.g., workplaces, sports events, military regions, airplanes) with defined stressors. Secondly, stress induction paradigms that are not restricted to laboratory settings would allow for more inclusiveness regarding participants. Older or home-bound participants (e.g., neuromuscular diseases, depression & anxiety), as well as participants who are living too far away from the laboratory, could be easily integrated into stress studies. Participants originating from different cultures could be easily reached at the same time, allowing for comparative cultural stress research.

Indeed, the emergence of the online TSST opens a whole new perspective on basic stress research. On the other hand, the procedure still relies on live interactions between participant, panel and experimenter. Thus, while being far more easily conductible and scalable than the laboratory TSST, performing large scale studies still remains a logistic challenge. The VR versions of the TSST theoretically could be applied without any experimenter contact but this would require participants to possess or obtain VR glasses, which is - at least currently - not realistic. In one study by Almazrouei et al. (2022) that was published shortly after the paper underlying this thesis, the authors adapt the Trier Mental Challenge Test Stress Protocol (Kirschbaum et al., 1992) for an automated online application, without any (direct or virtual) experimenter contact. This protocol is based on the well-known stress induction principles of uncontrollability and social-evaluative threat (Dickerson & Kemeny, 2004). Participants need to answer general knowledge and arithmetic questions under time pressure and receive negative feedback. In their study, 118 participants performed either the stress test or a comparable control test on an online platform. They measured the state anxiety and the subjectively perceived stress levels after the stress induction and could show significantly higher stress and anxiety levels in the stress group. While this can be seen as promising evidence for the possibilities of remote and independent stress induction, some limitations of the studies have to be discussed. First, participants' baseline stress levels have not been measured and thus, not been accounted for in the group comparison. Secondly, the stress levels are only measured with one visual analogue scale, overseeing potentially more psychological dimensions of stress (such as frustration and shame). Further, physiological stress indicators are still lacking and would need to be checked in follow-up studies. In general, this study tackles one important aspect of current stress research advancements - namely providing a remotely and independently feasible stress *induction* paradigm - that was also motivating the development of the DST. Having seen the potentials of a (scalable) remotely feasible stress induction paradigm, the following section argues (for using such a tool) to also include the possibility to collect (reliably annotated) non-verbal behavioral stress video data to the paradigm.

1.2 Stress Analysis from Video Data

Using video data for learning about the non-verbal behavioral properties of the stress response and/or developing stress prediction algorithms is a growing research area at the intersection of psychology and computer science. Researchers use videos of participants (e.g., recorded during laboratory experiments) that are annotated with some kind of stress level (e.g., perceived stress level questionnaires, cortisol indices, external annotators' impression of the participant's perceived stress), define and extract different features (e.g., facial movements, voice characteristics, pulse levels) and perform standard statistical tests (e.g., t-tests for group comparisons, regression) or train machine learning algorithms (e.g., Random Forest Classifier or neural networks). This way, conclusions about the non-verbal behavior of (more or less) stressed people can be drawn and used for predicting and feedbacking stress levels to a user, for example in monitoring drivers' stress (Gao et al., 2014).

To the current date, an extensive and comprehensive literature review on video based stress detection algorithms (including dataset properties, feature extraction, algorithms and results) is missing, although various works have been published in the past: For example, Giannakakis et al. (2017) developed an algorithm for detecting stress using facial cues extracted from videos in 2017 already. They could show that selected facial cues for discriminating between stress-induced and neutral states showed accuracy rates higher than 80% with different classification methods (e.g., support vector machine, AdaBoost). Additionally, they identified specific voluntary or semi-voluntary facial cues (e.g., eye-related, mouth activity and head movement features) that serve well for discriminating stressed from neutral states, thereby providing insights into behavioral correlates of the stress response.

Among other experiments, Aigrain et al. (2018) applied several support vector machines on behavioral features extracted from videos for classifying different stress dimensions and could also achieve F1 scores up to 0.8 for externally annotated stress levels but lower scores for predicting self-assessed (0.6) or physiological (0.79) stress levels. In their study, they also analyzed the predictive power of behavioral features and could show that body movement features were most informative across the different stress dimensions. Zhang et al. (2020) analyzed facial expressions and action motions in participant videos in an own dataset with a two-leveled stress detection network yielding detection accuracy of 85.42%.

Using the recorded speeches of 43 participants undergoing the Trier Social Stress Test, Baird et al. (2019) developed an algorithm to predict sequential measures of cortisol. They found highest correlations of ρ = .421 between a Support Vector Regression model predictions and cortisol measures. In another recent work, Baird et al. (2021) combined several TSST datasets and incorporated facial expressions as well as a multimodal approach to predict physiological (e.g., cortisol levels, heart rate, respiration) and externally annotated emotional (e.g., valence and arousal rating) stress. Using the same acoustic features with a deep neural network architecture (Long Short-Term Memory Recurrent Neural Network), they improved their results to correlations between model predictions and cortisol levels of ρ = .770. Other approaches implemented algorithms using the heart rate variability detected from facial videos (analyzing the subtle color changes of the skin due to heartbeat) to classify stress levels based on the State-Trait-Anxiety-Inventory questionnaire (Iuchi et al., 2020; Mitsuhashi et al., 2019). They reported accuracies of higher than 70% for different stress levels.

While this is not meant to be an exhaustive overview, the potentials of stress detection algorithms become clear. Preliminary results seem promising and systems are already tested for real-world applications such as driver monitoring (Gao et al., 2014) and student emotion inference during lectures (Tonguç & Ozaydın Ozkara, 2020). Nevertheless, there are reasons to investigate this more closely: First, it is hard to objectively evaluate and compare different works due to strongly varying methods and measures. Secondly, the applicability of such algorithms highly relies on the size, quality, representativeness and annotation of the development's underlying dataset which is not always considered. For example, a driver monitoring system trained on external annotated stress levels would not be able to predict physiological arousal or subjective feelings of stress. Lastly, most video datasets currently used for stress detection algorithm development are rather small and do not represent naturalistic settings (Roldán-Rojo et al., 2021). Thus, there is a need of collecting more video data, ideally from different participant cohorts, in different settings and with reliable stress level annotations. Additionally, the research community would highly benefit from an open science approach, where datasets, code and results are clearly communicated and set into context.

1.3 Objectives

The objective *within the accompanying publication* (Norden, Hofmann, et al., 2022) was to conceptualize a remotely and online feasible *stress induction* paradigm that can be conducted (with a smartphone) without any researcher contact. Additionally, we aimed for evaluating the feasibility of the tool in a large online study and validate the subjective stress induction against a suited control condition. *In this thesis*, I want to build upon this work and put the DST into the broader context of interdisciplinary psychological and computational stress research. I will tackle limitations discussed in the accompanying publication by describing current and future plans within the DST project. Therefore, I will present an on-going pre-registered follow-up laboratory validation study (Norden, Mayer, et al., 2022) and further technological developments of the DST.

Secondly, I want to discuss opportunities and limitations of the video recording capabilities implemented in the DST with respect to developing a new stress video dataset and stress detection algorithms. I will demonstrate how the definition and annotation of the stress response influences the results of algorithms developed with this data taking a published evaluation study of stress test video based machine learning algorithms into account (Norden, Wolf, et al., 2022).

Finally, this thesis serves as an invitation for interdisciplinary stress research to collaborate on a large scale and promote open science (e.g., making research protocols, data, and analysis and implementation code available for the community).

2 Methods

In this section, I will summarize the methods used for conceptualizing, developing and evaluating the DST as described in more detail in Norden, Hofmann, et al. (2022). Further, statistical methods that were applied for deriving results, which have not been published are included. Methodological details on the additional pre-registered follow-up laboratory study and the evaluation study on machine learning models for video-based stress detection are covered in the respective publications (Norden, Mayer, et al., 2022; Norden, Wolf, et al., 2022).

2.1 Concept & Development of the DST

The general idea of the DST is to provide a scalable (digital) way to perform stress studies independently and collect videos capturing stress related behavior of the participants. Thus, we designed the DST as an easy-to-use smartphone web application where subjects conduct the study (via internet) without the need for direct communication to the researchers or additional resources (i.e., wearables or native application downloads). To elicit an individual stress reaction, we implemented principles of classical stress paradigms suited to this digital setting. The DST framework also includes information on study background, privacy, consent and debriefing as well as several mood questionnaires which can be adapted for specific study purposes. To evaluate the stress induction potential of the DST, we developed a control version - the Control-DST (C-DST) - that resembles the general design and procedure of the DST but is changed with respect to stress induction elements. The code and prototypes of both web applications can be found on www.digitalstresstest.org/code. Screen recordings of the versions used within the present study can be found in the Multimedia Appendices of the accompanying publication (Norden, Hofmann, et al., 2022).

2.1.1 Digital Stress Induction Paradigm

The Digital Stress Test paradigm comprises a mental arithmetic part and a verbal answering task. Both tasks and the underlying app framework are enhanced with several stress-inducing framings and other functional or design elements to elicit feelings of uncontrollability and social-evaluative threat (Dickerson & Kemeny, 2004). First, the whole procedure is introduced as a “cognitive-verbal performance test” where the participants are video recorded by the front camera, and their behavior gets analyzed. The first task is then framed as a simple calculation task that a faked comparison group would have solved with an average of 75% correct answers.

In the calculation task (“Math-Task”) participants have to solve simple calculation tasks (i.e., addition, subtraction, multiplication, division of two numbers ranging from 1 to 99 with solutions ranging from 1 to 99) on a number field within a given time limit. Negative visual (e.g., screen turns red) and verbal feedback is provided for wrong or too slow inputs. No positive feedback is provided for correct responses but the next task is displayed immediately. Participants do not get informed on the duration of the whole task. The difficulty of the tasks is adaptively changing depending on the current performance of the participant. Therefore, the time limit for answering the tasks is shortened by 10% when the participant answers a series of three consecutive tasks correctly. At the same time, the order of the input field number changes randomly and stays like this for the following four tasks. The limit is extended by 10% when the participant fails to correctly respond to the tasks for three times in a row. To motivate on-going participation, a prompt that indicates the relevance of the participation is displayed when the participant does not give any input for five consecutive tasks. Additionally, the following task is programmed to be an easily solvable summation task. During the whole Math-Task, the participants see a bar chart comparing their current percentage of correct responses with the apparent average of 75% for the faked comparison group. Due to the implemented difficulty adaptation algorithm, the participants will perform worse than the comparison group which reminds the participant of failing in the cognitive performance task. To remind the participants that they’re being recorded and apparently analyzed, the video recorded by the smartphone front-camera is directly displayed in the upper half of the screen for the whole Math-Task. After 1.5 minutes the Math-Task automatically stops and the participants see their final performance compared with the fabricated performance of their comparison group.

The verbal answering task (“Speech-Task”) is framed to assess verbal skills within presentation-like scenarios similar to job-interview situations. The Speech-Task contains three inconvenient questions (Fehlner et al., 2020) but participants are not informed on how many and which questions will follow. For each question, the participants are instructed to prepare convincing verbal responses for ten seconds. They then have to present their response in front of the camera for twenty seconds. Preparation and presentation times are displayed through a countdown bar. The smartphone front camera video recording is displayed in the upper half of the screen. During presentation periods only, the whole screen blinks red and a voice visualization is displayed. If no noise input signal is recorded for one second, the participant gets reminded to keep on talking.

2.1.2 Control - DST

To ensure a comparable baseline, the Control - DST (C-DST) procedure exactly matches the DST procedure up to the first mood evaluation questionnaire. Afterwards the general structure and design are similar but the framing and the task details differ. The participants get informed that they are part of the control group and that no video recording takes place. There is no framing of a comparison to other participants.

The C-DST Math-Task is implemented to be much easier: Only summation tasks are displayed and the initial time limit is longer. The time limit is extended by 10% as soon as the participant answers one task incorrectly or not at all. The time limit is shortened by 10% when the participant answers a series of four consecutive tasks correctly. The order of the input number field does not change. Layout and feedback are chosen to be more encouraging: The screen color changes to green for correct answers and does not change for wrong answers. The time limit is displayed using a green progress bar. Neither a faked comparison to other participants nor any live recordings of the front camera are shown. The C-DST Speech-Task is not introduced as an assessment of verbal skills. The three answering scenarios include neutral topics. Neutral images are displayed in the upper half of the screen. The colors are chosen to be more calming. There is no distraction through a red blinking background.

2.1.3 Technological Framework & Security

The system architecture of the applications is shown in **Figure 1** and explained in more detail in the accompanying publication. Both applications were developed as single-page web applications with the JavaScript framework React.js. The applications are hosted on a university server using the open-source study management system JATOS (Lange et al., 2015). For the purpose of the published study, the storage of video data was disabled and only age and gender inputs, meta data (i.e., study progress, durations) , Math Task inputs and sound levels recorded during the Speech Task were collected. Security measures included using NginX and ensuring Secure Sockets Layer (SSL) encryption and restrict access with Hyper Text Transfer Protocol Secure (HTTPS). For securely saving potentially identifiable video data we also implemented a secure copy protocol

(SCP) that automatically transfers data from the publicly reachable university server to a separate storage server.

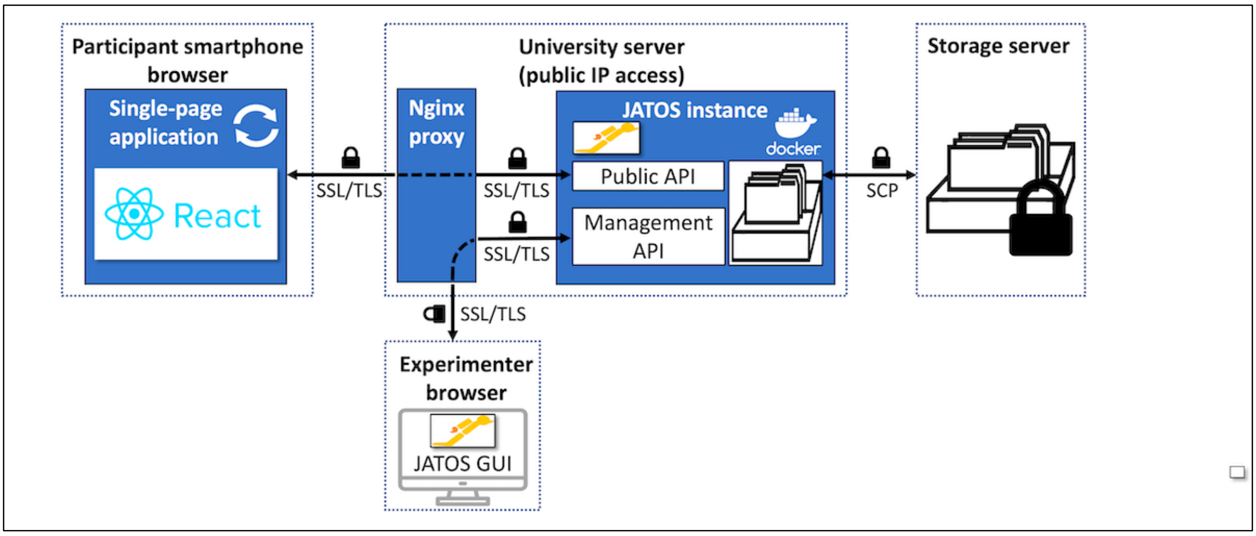


Figure 1. System architecture of the test applications (taken from Norden, Hofmann, et al., 2022).

2.2 Online Study

To evaluate the feasibility and the stress induction potential of the DST, we conducted a large online study comparing affective responses of DST and C-DST participants indicated in different questionnaires.

2.2.1 Participants and Recruitment

Participants were recruited via publication of the study participation link in several mailing lists, social networks, podcasts and websites. Data collected for this study reflect a period of two weeks, from February 10 to February 24, 2021. During this time, 547 participants started the study (see **Figure 2**). Participants with incomplete tests (n=229), previous self-reported knowledge of the framing (n=13), self-reported usability issues (n=19), or unrealistic study duration (n=2) were excluded.

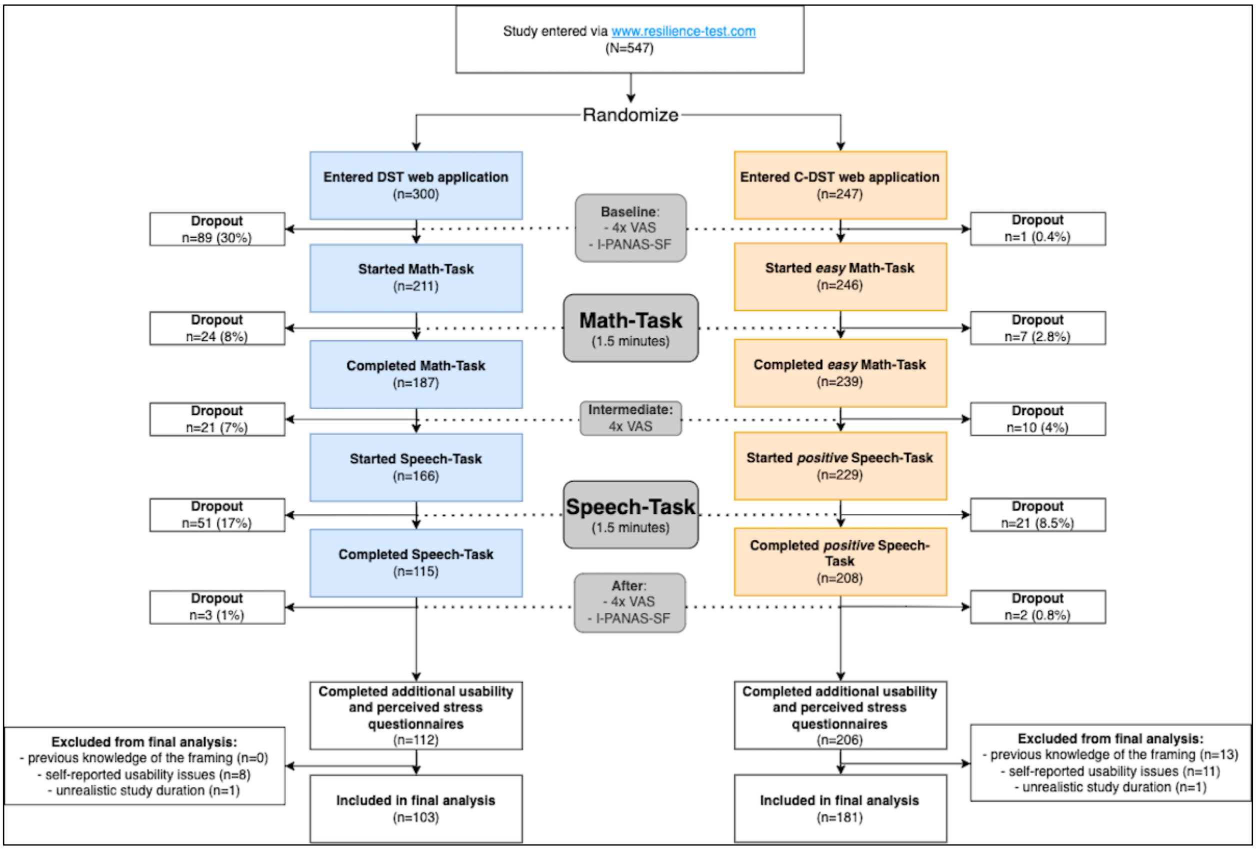


Figure 2. Online study design including number of participants and dropouts at different stages of the study (taken from Norden, Hofmann, et al., 2022).

2.2.2 Procedure

Participants were randomly forwarded to one of the two web applications after clicking the study link. Before starting the actual paradigm with the first task, participants of both applications were informed on study background, procedure and privacy aspects in the same way. Participants filled out built-in affective state assessment questionnaires - the Visual Analogue Scales (VAS) and the international short form of the Positive And Negative Affect Scale (i-PANAS SF) - before, during and after the two tasks. Debriefing information and a link to follow-up questionnaires on an external website were displayed at the end of the tests. The study was approved by the Ethics Committee of University of Potsdam (application 33/2020) and was conducted in accordance with the General Data Protection Regulation.

2.2.3 Statistical Analyses

Data analyses were performed using Python 3.7. The code for preprocessing and analyzing the data has been published on a public repository which can be found at https://github.com/mno-93/DST-validation-study-analysis. First, distributions of the affective response data collected using VAS and PANAS were assessed for normality and homogeneity of variances using Shapiro-Wilk and Levene tests. Afterwards, we used two separate mixed-model analyses of variance (ANOVA) for repeated measurements with the factor TIME (baseline and post for PANAS; baseline, intermediate, post for VAS) and the between-participant factor GROUP (DST vs. C-DST). We used Greenhouse-Geisser corrections for the degrees of freedom where the sphericity assumption was not fulfilled. Because of different group sizes and nonhomogeneity of variances, post hoc comparisons were performed using Welch's t-test. Additionally, mean VAS scores for the subjective stress experience regarding different aspects of the web applications were calculated and ranked.

As a further evaluation, affective responses of DST participants in this study were compared to those of participants conducting the TSST in four previous studies (Herten et al., 2016; Herten, Otto, et al., 2017; Herten, Pomrehn, et al., 2017; Wolf et al., 2015). Therefore, we analyzed the normalized PANAS subscores (i.e., positive and negative affect) in a two-step meta-analysis. For each TSST study, a paired t-test on the normalized pre- and post PANAS subscores was performed and the standardized effect size calculated. We then combined the individual TSST study effect sizes by assigning weights based on the inverse of the change score variance (Borenstein et al., 2009) and compared it to the standardized effect size observed in DST participants.

As part of the online study but not published in Norden, Hofmann, et al. (2022), we compared the individual performances during the Math-Task and the speaking participation between DST and C-DST using Mann-Whitney-U tests. As a measure of performance during the Math-Task, we used the percentages of a participant's correct answers. To assess a participant's participation in the Speech-Task, we calculated the percentage of sound input reaching a pre-set threshold during the presentations. Additionally, we analyzed the relationships of performance during the Math-Task, speaking activity during the Speech- Task and the subjective stress experience using Spearman correlations.

3 Results

In this section most important results of the accompanying publication are summarized. This includes general information on the online study and the affective response evaluation, described in detail in Norden, Hofmann, et al. (2022). Additionally, unpublished analysis results regarding the context in which the users performed the tasks, and the relationship of performance during the Math-Task and speaking activity during the Speech-Task to the subjective stress experience are presented.

3.1 Participants and Study Procedure

Number of participants and dropouts at different stages of the study are illustrated in **Figure 2**. From 547 participants that started the study, 103 individuals (50 men, 52 women, 1 other) completed the DST and 181 individuals (83 men, 96 women, 2 other) the C-DST. DST participants finished the procedure in 7.69 minutes (SD 1.35 minutes) and C-DST participants took 6.53 minutes on average (SD 1.05 minutes). Overall 229 individuals did not finish the study. Within the C-DST group, 41/247 (16.6%) participants dropped out whereas 188/300 (62.7%) participants dropped out after starting the DST. Most DST dropouts (89/188) took place before starting the Math-Task.

3.2 Affective Response Evaluation

The affective responses indicated by the participants at different time points during the procedure are illustrated in **Figure 3** and **4**.

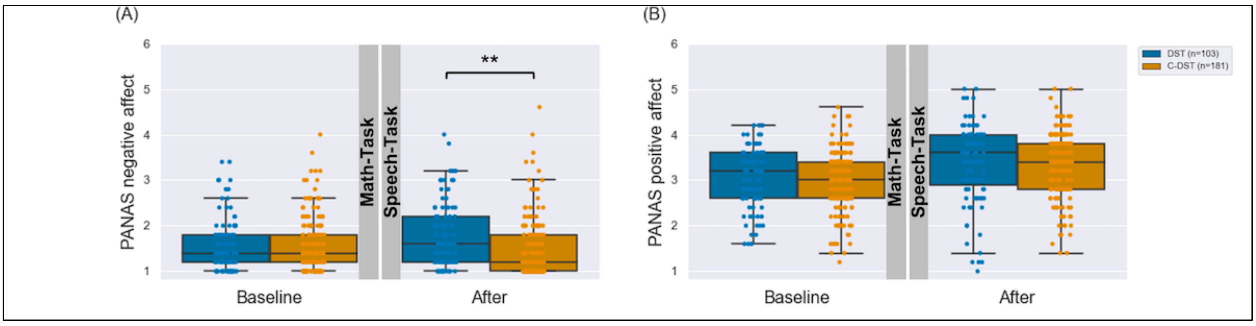


Figure 3. Digital Stress Test (blue) and Control - Digital Stress Test (orange) participants' indicated negative (A) and positive (B) affect in the Positive and Negative Affect Schedule (PANAS) subscales before and after stress induction (taken from Norden, Hofmann, et al., 2022).

Using mixed-model ANOVA for the PANAS negative affect subscale measurements, we found a significant main effect for the factor GROUP ($F_{1,282}$ = 5.83; p = .02; η_p^2 = 0.02) and a significant GROUP x TIME interaction effect ($F_{1,282}$ = 31.37; p < .001; η_p^2 = 0.10). Participants' overall reported negative affect was higher in the DST (mean 1.70, SD 0.55) than C-DST (mean 1.54, SD 0.57) group (p < .001). In the baseline measurements, participants' indicated negative affect was not significantly different between DST (mean 1.57, SD 0.56) and C-DST (mean 1.58, SD 0.58) group (p = .99) whereas negative affect in the post-test assessment was significantly higher in the DST (mean 1.84, SD 0.7) than C-DST (mean 1.49, SD 0.64) group (p < .001). The mixed-model ANOVA for PANAS positive affect scale yielded a significant main effect for the factor TIME ($F_{1,282}$ = 0.43; p < .001; η_p^2=0.002) but neither for the factor GROUP (p = .40) nor for the interaction GROUP x TIME (p = .51). Overall, perceived positive affect was significantly higher after (mean 3.34, SD 0.76) than before (mean 3.02, SD 0.65) conducting the tests (p < .001).

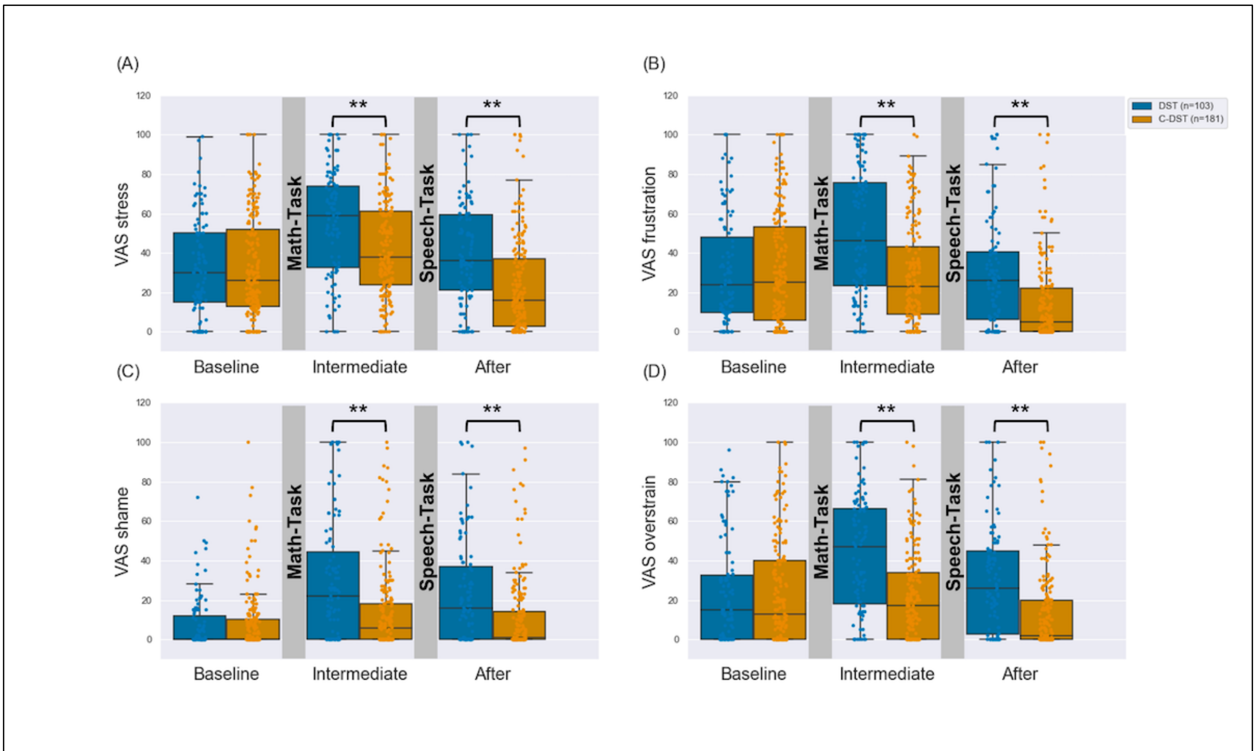


Figure 4. Digital Stress Test (blue) and Control – Digital Stress Test (orange) participants' indicated affect in the four visual analogue scales at different time points (taken from Norden, Hofmann, et al., 2022)..

Separate mixed-model ANOVA for the VAS related to stress showed significant main effects for the factors of GROUP ($F_{1,282}$ = 14.42 ; p < .001; η_p^2 = 0.05) and TIME ($F_{2,564}$ = 75.11; p < .001; η_p^2 = 0.21), and a significant interaction between these factors ($F_{2,564}$ =

14.28; $p < .001$; $\eta_p^2 = 0.05$). Overall, reported stress responses were higher for DST (mean 42.39, SD 20.87) than for C-DST (mean 32.79, SD 20.27) participants ($p < .001$). Perceived stress levels of DST participants significantly increased over the course of the Math-Task, from a baseline of 32.92 +- 25.48 to an intermediate level of 46.76 +- 26.01 ($p < .001$). It then significantly decreased over the Speech-Task, from the intermediate level of 46.76 +- to a level of 29.15 +- 25.96 after the procedure ($p < .001$). Similar patterns were found for three other stress-related attributes (frustration, shame, and overstrain) when conducting separate mixed-model ANOVA and post-hoc tests.

Further, we compared the stress-inducing potential of the DST with the results of four previously conducted TSST studies (**Table 1**) in a two-step meta analysis. Participants reported a significantly higher negative affect after conducting the tests in all but one TSST study. The combined effect size for the change in negative affect of the TSST (combined $d_z = 0.667$) studies was slightly higher than the effect size for the change in negative affect of DST participants ($d_z = 0.427$) in this study. Participants indicated a significantly higher positive affect after conducting the DST whereas there was no evidence for significant increases in TSST participants' positive affect.

Table 1: Overview of studies used for meta-analytical comparison of the Digital Stress Test with the Trier Social Stress Test (TSST). Adapted from Norden, Hofmann, et al., (2022).

Study, PANAS[a] subscale	Score before, mean (SD)	Score after, mean (SD)	Change, mean (SD)	Welch t-test	
				P value	d_z
Digital Stress Test (N=103) – Norden, Hofmann, et al. (2022)					
NA[b]	1.57 (0.56)	1.84 (0.70)	0.27 (0.61)	**<.001**	0.427
PA[c]	3.08 (0.65)	3.37 (0.85)	0.29 (0.61)	**<.001**	0.382
TSST (N=26) - Herten, Pomrehn, et al. (2017)					
NA	1.28 (0.38)	1.82 (0.66)	0.55 (0.68)	**<.001**	1.015
PA	2.95 (0.45)	2.94 (0.59)	−0.01 (0.51)	.91	0.022
TSST (N=26) - Herten, Otto, et al. (2017)					
NA	1.33 (0.32)	1.66 (0.57)	0.32 (0.44)	**<.001**	0.694
PA	2.95 (0.56)	2.97 (0.64)	0.02 (0.45)	.86	0.026
TSST (N=20) - Herten et al. (2016)					
NA	1.36 (0.33)	1.51 (0.71)	0.16 (0.68)	.32	0.281
PA	2.75 (0.42)	3.02 (0.96)	0.27 (0.82)	.16	0.363
TSST (N=50) - Wolf et al. (2015)					
NA	1.43 (0.56)	1.85 (0.72)	0.42 (0.53)	**<.001**	0.655
PA	3.02 (0.57)	2.88 (0.68)	−0.14 (0.47)	.04	0.221

[a]PANAS – Positive And Negative Affect Schedule

[b]NA – Negative Affect

[c]PA – Positive Affect

3.3 Context, User Experience and Stress Element Analysis

After finishing the web application procedure, participants were asked to answer general questions on the situation during which they performed the test, and on usability and functionality of the web apps. Further, we asked them to rate their subjective stress perception regarding specific elements within the procedure. Participants conducted the task in 10 different nations, at varying times of the day (e.g. 21% of the participants performed the study in the morning, 34% in the evening), diverse locations (95% at home, 3% at work, 2% on the way), different postures (80% sitting, 14% lying, 6% standing), with (13%) and without (87%) presence of other people and in silent (67%) and non silent (33%) environments. Most participants (92.6%) did not report any usability problems. Participants, that reported limited readability of the texts or could not properly use the VAS implementation, were excluded from the affective response evaluation. Other usability issues included scrolling issues within the Math-Task or problems regarding the responsiveness of the layout for different smartphones and browsers. The results of the element-specific subjective stress perception analysis are displayed in **Table 2**. DST participants indicated that the Math-Task was the most stressful part during the procedure, followed by the Speech-Task and the framings in the beginning. For both tasks, the participants indicated that time pressure was the most stressful aspect. Being recorded through the front camera was not perceived as that stressful. Randomly swapping the input field after three consecutive correct task responses seemed to induce a high level of perceived stress.

In addition to the published results, we analyzed the influence of the personal performance during the Math-Task and the speaking participation during the Speech-Task on the subjective stress levels. First, we compared the individual performances during the Math-Task as well as the speaking participation between DST and C-DST. Participants of the DST achieved significantly lower performances (*Mdn* = 46.67%) than participants performing the C-DST (*Mdn* = 82.14%), $U = 90.00$, $p < 0.001$. Participants of the DST participated significantly more in the Speech-Task (*Mdn* = 11.11%) than C-DST participants (*Mdn* = 6.00 %) based on the recorded sound inputs, $U = 9048.00$, $p = 0.04$. Secondly, we analyzed the relationship between the DST-participants' performance or speaking participation and their changes on the VAS stress scales for the two tasks (i.e., pre-to-intermediate for Math-Task performance, intermediate-to-post for Speech-Task participation) performing Pearson correlation analysis. There was no evidence for significant associations between the performance in the Math-Task and the change of indicated stress levels ($r(101) = -.15$, $p = .13$) nor between the speaking participation during the Speech-Task and the changes of indicated stress levels ($r(101) = -.15$, $p = .13$).

Table 2: Elements of the Digital Stress Test (DST) and perceived stress levels (adjusted from Norden et al., 2022).

Stress Element in the DST	VAS[a] Mean ± SD
Framing	n/a[b]
Participation in a performance test	65.4 ± 23.8
Behavior analysis through algorithm	40.6 ± 27.5
Math-Task (overall)	77.9 ± 18.4
Time limit	88.8 ± 13.5
Random input field swap	88.6 ± 15.6
Feedback after every calculation task	67.3 ± 24.0
Task difficulty	62.8 ± 26.3
Live comparison to other participants	59.8 ± 28.5
Personal results / performance	58.0 ± 29.0
Front camera activation	39.7 ± 28.6
Speech-Task (overall)	46.2 ± 23.3
Preparation periods	45.5 ± 25.8
Time limits	45.1 ± 27.3
Questions	43.6 ± 23.8
Front camera activation	37.1 ± 25.9
Audio visualization of the voice	30.2 ± 24.2

[a]VAS – Visual Analogue Scale

[b]n/a – Not applicable

4 Discussion

In the accompanying publication (Norden, Hofmann, et al., 2022), we assessed the viability of a smartphone-based, fully digital stress paradigm (DST) to induce psychosocial stress outside of a laboratory setting. The DST was compared to a digital control condition (C-DST) in a large online study and the participants' affect changes were set in context with those previously seen in the Trier Social Stress Test (TSST). It was the first study to evaluate the stress responses of an experimenter-independent, automated paradigm without any human-human interaction. In the following, I will critically discuss the presented results on the basis of the accompanying publication including interpretations of the results with respect to related work, potentials of the DST, and limitations. Further, I include additional information on further developments of the DST, additional validation and machine learning studies (Norden, Mayer, et al., 2022; Norden, Wolf, et al., 2022), and future plans within the project.

4.1 Principal Findings

Results showed that the DST induced significantly higher levels of perceived stress and negative affect compared to the control condition, with participants also reporting increases in related affects such as frustration, shame, and overstrain. Remarkably, the DST participants' reported increases in negative affect were not only greater than those indicated by C-DST participants but comparable to those seen in previous TSST studies, as shown by calculated effect sizes.

With this study, we provided initial evidence that subjective stress can be induced outside the lab in a standardized way without requiring further equipment or any interaction with researchers. This is in line with results obtained with other innovative and less controlled stress induction paradigms like different virtual reality or online versions of the TSST. A recent meta-analysis on VR-TSSTs concluded that the effect sizes of the stress reactivity were comparable to those calculated in traditional TSST (Helminen et al., 2021) and Gunnar et al. (2021) reported effect sizes in the range of traditional TSST for their online version. On the other hand, some studies reported successful stress inductions with weaker stress responses (Montero-López et al., 2016; Shiban et al., 2016; Turner-Cobb et al., 2019), attributing this mainly to a lack of social evaluation stress when the protocol does not include direct human-human interactions.

The study that was evaluating the stress responses of a fully automated online paradigm shortly after the results of the DST were published (Almazrouei et al., 2022), reported successful subjective stress inductions. While the virtual reality and online versions of the TSST have already been shown to also induce physiological changes, the fully automated online paradigms' stress induction potential still needs to be further evaluated.

Interestingly, the Math-Task seemed to evoke a stronger stress response than the Speech-Task whereas in previous studies especially public speaking parts have been shown to induce higher stress levels in participants (Dickerson & Kemeny, 2004; Westenberg et al., 2009). This might be due to the lack of direct social interaction or human evaluation in comparison to live experiments. Additionally, despite the automated feedback to continue speaking, it could be that participants did just not follow the instructions. They were informed that their videos would be analyzed but knew that the videos were not permanently stored or watched by human evaluators. We already received an ethics approval for an additional online study during which the videos will be stored and participants informed that their behavior will be evaluated. Furthermore, we already implemented a faked speech quality feedback component into the current version (www.digitalstresstest.org) which intends to put an additional social evaluative stressor to the task and remind the participants of failing. We hypothesize that the stress perceived over the Speech Task will be stronger in the upcoming study than in the published one.

The potential of performing automated online studies became clear in this study already. The number, composition and study contexts of participants exceeded those observed in previous stress induction studies by far (Campbell & Ehlert, 2012). On the other hand, the sample in this study was still biased with respect to the age and educational background which points to the influence of advertising channels. This study was widely spread through a wide-spread German political podcast (Banse & Buermeyer, 2021) and university mailing lists. Additionally, even if the internet is now used across all ages, younger people tend to spend more time with it (Bucur et al., 1999; Hargittai et al., 2019) which further exacerbates the age bias. Thus, there are several points to consider when interpreting the results of this study:

First, the stress reactivity can differ across ages and backgrounds (Gotthardt et al., 1995; Neupert et al., 2006). It could theoretically be that the DST might be able to induce subjective stress in a young, higher-level educated population but fails to do so in other cohorts and hence, could not be applied for general stress studies. Secondly, as the purpose of the DST is also to collect a multimodal stress dataset for developing video-based stress detection systems, these biases would be propagated to the algorithms when not handled carefully. While we do not expect big differences in terms of stress induction, we plan to broaden the study advertisement strategies and specifically account for age, culture and educational backgrounds in the upcoming online study. We aim to ensure a representative and unbiased study sample for creating the video dataset and the respective algorithms.

Two challenges in conducting fully automated online studies are the lower barrier for dropping out and the lack of control. We kept the procedure short and implemented several mechanisms to remind the participant to follow the instructions and also logged the user progress. Nevertheless, it’s hardly possible to account for confounding factors such as situational aspects or technological problems. We wish to show the potential of the DST to be applicable in various contexts and thus, gathering vastly more diverse data. For the upcoming online study, we plan to include the contexts into the stress reactivity analysis and also evaluate the video and sound recordings more specifically.

As discussed in Norden, Hofmann, et al. (2022), a larger portion of DST participants dropped out in the beginning which might be due to the video recording framing or technological problems with the camera. Even if the participants knew that the recordings would not be saved, privacy concerns might have led to the dropouts (Shore, 2022).

High standards of data protection, a clear communication to the participants, an active consent and the possibility to withdraw with complete data deletion at any time are crucial, especially when dealing with personal and potentially identifiable data such as videos (American Psychological Association, 2017). Therefore, we already worked on the previous DST and C-DST applications and included several aspects after the initial publication: First, participants get briefly informed within the app as in the previous version but now need to download an additional information, consent and data protection sheet. Afterwards they have to actively confirm that they’ve read and understood the information before they can continue with the procedure. Secondly, we implemented a clearly visible “cancel button” that allows for cancellation at any time. When pressing the button, participants are asked to confirm their cancellation and can decide if data obtained so far might still be used for the study. While the default is that all data is deleted, they can still decide that the data (questionnaires, age, gender, meta data) including or excluding videos will be saved. Afterwards, they get directly linked to the debriefing page. If participants cancel the study by closing the browser, no data is permanently stored. The participants have to actively confirm the submission of data at the end of the study. An

individual code is displayed to the participants and they get informed that they can contact the study team for requesting insight into or deletion of their data.
The whole procedure was already reviewed by the ethics committee and data protection office of Bielefeld University.

4.2 Limitations

The study presented in Norden, Hofmann et al. (2022) clearly illustrated the feasibility of an automated online stress induction procedure. While the participants' subjective stress significantly increased throughout the procedure, the influence on physiological stress reactivity is still not clear. Addressing subjectively perceived stress during daily life plays an important role for the individual quality of life and mental well-being. Physiological stress responses have been clearly associated with long-term consequences for cardiac, metabolic and mental health (Yaribeygi et al., 2017). Therefore, targeting both physiological and subjective stress reactivity, would be of high value for preventive care. Even if there are some studies indicating a correlation of subjective and physiological stress levels (Rimmele et al., 2007; Schlotz et al., 2008), the results are not conclusive. Thus, the DST should be validated separately for the induction of acute physiological stress responses. This would open up even more possibilities for the stress research community due to the easy and scalable applicability in different settings, cohorts and contexts (Kirschbaum, 2021).
We already prepared, piloted and started a laboratory study for the validation of the DST. The protocol for this study has been published on an open science online platform (Norden, Mayer, et al., 2022). The procedure of the study is summarized in **Figure 5**. We compare subjective, observer-rated and physiological stress responses of participants undergoing the DST and TSST on separate days in a randomized within-participant design. Next to measuring several aspects of the physiological stress reaction, including parameters of the HPA-axis (i.e., salivary cortisol), the sympathetic-adreno-medullar (i.e., salivary alpha amylase) and related circulatory system (i.e., heart rate, blood pressure, breathing rate), participants get filmed during the experiments. Further we already prepared and received an ethical approval for an additional between-subject study with a similar design comparing stress responses to the DST and C-DST. In this study, also the influence of the DST on short-term cognitive functioning will be analyzed. The planned procedure is outlined in **Figure 6**.

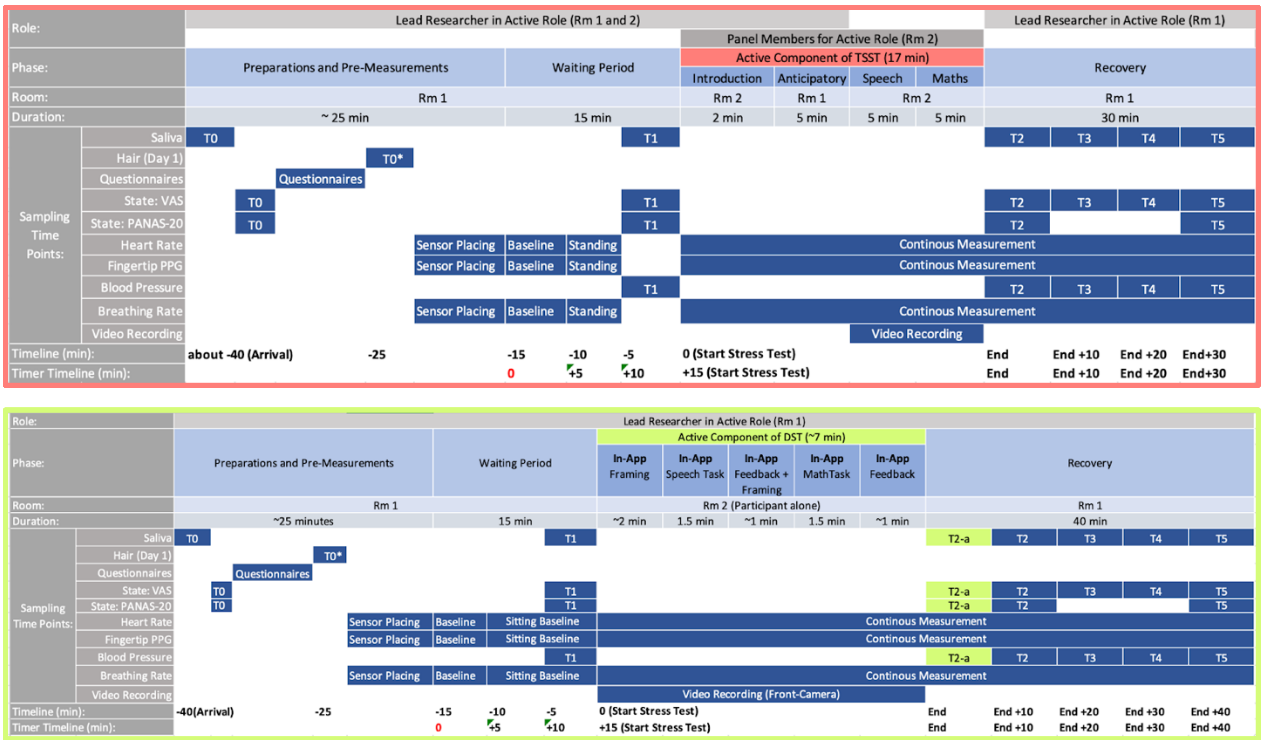


Figure 5. Detailed study design for the pre-registered follow-up validation study of the DST taken from the updated pre-registration (Norden, Mayer, et al., 2022). Upper and lower row display the study procedure including general timeline, measurements and measurement time points on TSST and DST day respectively.

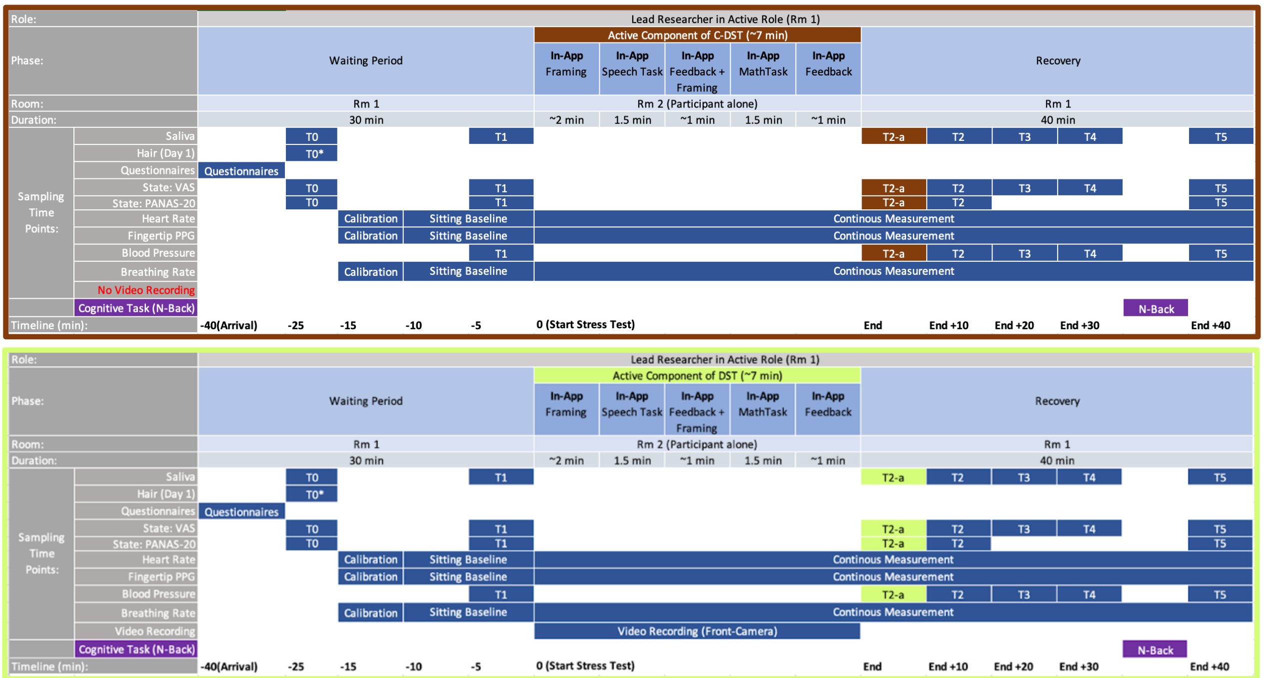


Figure 6. Unpublished study design for the additionally planned follow-up validation study of the DST. Upper and lower row display the study procedure including general timeline, measurements and measurement time points of the C-DST and DST experiments in a between-participant design respectively.

We expect significant increases in stress indices for the DST and TSST but not C-DST experiments. We plan to use the videos of all experiments to develop and improve stress detection algorithms.

The ability to gather stress-related non-verbal behavior in videos online has not been evaluated in the published study as the purpose of this study was evaluating the stress induction potential of the DST. Further, the DST was applied on smartphones and tablets exclusively. While this allowed for showing the potential of conducting stress experiments without further equipment, we now also developed a responsive desktop version. Therefore, workplace related experiments can be performed and less movement-sensitive videos recorded. Evaluating the quality of the online gathered video recordings and evaluating standard feature extraction methods will be one of the first steps in the planned online study.

4.3 Future Research

Some directions of future research with the DST have already been outlined. These include the improvement of the stress induction procedure by evaluating newer versions of the DST, validating the DST for physiological stress induction in two laboratory studies, and the additional online study for collecting a large and representative stress video dataset. A main focus for the next project steps will be using the videos of the currently on-going laboratory studies and future videos from the outlined online study for developing multimodal stress detection algorithms.

As published in Norden, Wolf, et al. (2022), we already developed baseline algorithms for detecting different dimensions of stress in videos of the Trier Social Stress Test. We analyzed videos of 40 TSST participants regarding subjectively perceived, live observer-rated, post-hoc video-annotated and biological stress levels and evaluated how the choice of the stress label is influencing classical machine learning models' performances. We could show that the machine learning models trained on different stress assessments using facial or voice characteristics perform differently and should be interpreted and applied accordingly. In line with previous works on video-based stress detection methods, the necessity of bigger and representative datasets for developing such models became clear. Therefore, we aim to leverage the scalability and easy application of the DST to collect a video dataset of participants with different ages, languages, educations and cultural backgrounds, in various contexts, and combine it with measurements of different dimensions of stress.

Next to developing and implementing the stress detection algorithms with the collected data, we aim to share the DST as an easy-to-use and scalable research tool across the community. After publishing the DST, we already received several requests for using the DST in different projects. While we intentionally published the code for interested researchers to set up and adapt the web applications for their own purposes, we also support other groups regarding ethical considerations, development and secure data collection. Currently, we are conducting a project with the University Würzburg and we expect more collaborations in the near future.

On the one hand, the DST might well serve as the induction tool presented in this thesis and therefore be included in any kind of study that wishes to implement a stress condition. On the other hand, one vision for the DST combined with the respective future stress detection algorithms is also to make it an objective stress reactivity and resilience test. Such future DST might help to recognize stressful times even when users are not aware of it and support preventive strategies. It might then also assist individuals to learn what triggers their stress reaction and how to cope with it best.

4.4 Ethical Considerations

As many of our studies involve stress induction procedures and/or the collection of personal and highly sensitive (health related) data, I consider ethical thinking as one core value for this research. Firstly, all studies described in this thesis, regardless of whether they are already finished, currently on-going or outlined, have been thoroughly discussed and prepared together with the ethics committee of University of Potsdam (applications 33/2020, 90/2020), Ruhr University Bochum or Bielefeld University (applications 2021-209, 2022-024). Regarding privacy and data protection, we consulted the competence center for research data of Bielefeld University and implemented several security measures with the help of the technical support team of the Faculty of Technology. Still, working in the fast-paced and constantly changing field of computer science, particularly when handling sensitive participant data related to health, might lead to potentially harmful consequences. While the potentials and applications of video-based stress detection methods are clear, these algorithms could potentially be used to analyze videos without prior consent or knowledge of individuals, for example in the case of surveillance.

Sensitive mental health information might be misused to discriminate against individuals, such as at the workplace or in the case of insurance coverage.

The accuracy of such methods is impacted by a range of factors, including cultural differences (Mehrabi et al., 2021) and the correctness of the underlying ground truths (Aigrain et al., 2018; Norden, Wolf, et al., 2022) in the development dataset. It is essential to ensure that these technologies are accurate, reliable, and sensitive for different groups of users. Therefore, the algorithms should be used with caution and (re-)evaluated for specific use cases before applying them.

5 Conclusion

In this project, we not only conceptualized and developed first prototypes of a standardized, fully automated and remotely feasible stress induction paradigm and a control version but also evaluated their potential in a large online study. We could show that the web applications can be used on smartphones without experimenter contact or needing further equipment and still induce psychosocial stress. The applications and their code have been made publicly available and researchers are invited to use this novel approach of conducting basic and clinical stress research.

Going beyond the publication, current and future improvements of the prototypes as well as further laboratory validation studies that follow up on the published work have been described. Moreover, in this thesis, the DST is set into the broader context of developing video-bas ed stress detection methods. Therefore, baseline machine learning approaches tested in an additional study are considered and ethical implications discussed. With the tools and the research described in this thesis, I aim to contribute to an unbiased, ethically thinking and open access way of conducting stress research in the intersection of psychology and computer science.

References

Aigrain, J., Spodenkiewicz, M., Dubuisson, S., Detyniecki, M., Cohen, D., & Chetouani, M. (2018). Multimodal Stress Detection from Multiple Assessments. *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*, *9*(4), 491–506. https://doi.org/10.1109/TAFFC.2016.2631594

Almazrouei, M. A., Morgan, R. M., & Dror, I. E. (2022). A method to induce stress in human subjects in online research environments. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01915-3

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. B., Alzakari, N., Abou Elwafa, A., & Kurdi, H. (2021). Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Applied Sciences*, *11*(2), Article 2. https://doi.org/10.3390/app11020796

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. Https://Www.Apa.Org. https://www.apa.org/ethics/code

Baird, A., Amiriparian, S., Cummins, N., Sturmbauer, S., Janson, J., Messner, E.-M., Baumeister, H., Rohleder, N., & Schuller, B. W. (2019). Using Speech to Predict Sequentially Measured Cortisol Levels During a Trier Social Stress Test. *Interspeech 2019*, 534–538. https://doi.org/10.21437/Interspeech.2019-1352

Baird, A., Triantafyllopoulos, A., Zänkert, S., Ottl, S., Christ, L., Stappen, L., Konzok, J., Sturmbauer, S., Meßner, E.-M., Kudielka, B. M., Rohleder, N., Baumeister, H., & Schuller, B. W. (2021). An Evaluation of Speech-Based Recognition of Emotional and Physiological Markers of Stress. *Frontiers in Computer Science*, *3*. https://www.frontiersin.org/article/10.3389/fcomp.2021.750284

Banse, P., & Buermeyer, U. (n.d.). *Lage der Nation* (No. 228). Retrieved April 19, 2023, from https://lagedernation.org/

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley and Sons. https://doi.org/10.1002/9780470743386

Bucur, A., Renold, C., & Henke, M. (1999). How Do Older Netcitizens Compare with Their Younger Counterparts? *CyberPsychology & Behavior*, *2*(6), 505–513. https://doi.org/10.1089/cpb.1999.2.505

Campbell, J., & Ehlert, U. (2012). Acute psychosocial stress: Does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*, *37*(8), 1111–1134. https://doi.org/10.1016/j.psyneuen.2011.12.010

Dedovic, K., Renwick, R., Mahani, N. K., Engert, V., Lupien, S. J., & Pruessner, J. C. (2005). The Montreal Imaging Stress Task: Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry and Neuroscience*, *30*(5), 319–325.

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

Eagle, D. E., Rash, J. A., Tice, L., & Proeschold-Bell, R. J. (2021). Evaluation of a remote, internet-delivered version of the Trier Social Stress Test. *International Journal of*

Psychophysiology: Official Journal of the International Organization of Psychophysiology, *165*, 137–144. https://doi.org/10.1016/j.ijpsycho.2021.03.009

Eurofound and EU-OSHA. (2014). *Psychosocial Risks in Europe—Prevalence and Strategies for Prevention*. https://doi.org/10.2806/70971

Fehlner, P., Bilek, E., Harneit, A., Böhringer, A., Moessnang, C., Meyer-Lindenberg, A., & Tost, H. (2020). Neural responses to social evaluative threat in the absence of negative investigator feedback and provoked performance failures. *Human Brain Mapping*, *41*(8), 2092–2103. https://doi.org/10.1002/hbm.24932

Gao, H., Yüce, A., & Thiran, J.-P. (2014). Detecting emotional stress from facial expressions for driving safety. *2014 IEEE International Conference on Image Processing (ICIP)*, 5961–5965. https://doi.org/10.1109/ICIP.2014.7026203

Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P., Marias, K., & Tsiknakis, M. (2017). Stress and anxiety detection using facial cues from videos. *BIOMEDICAL SIGNAL PROCESSING AND CONTROL*, *31*, 89–101. https://doi.org/10.1016/j.bspc.2016.06.020

Gotthardt, U., Schweiger, U., Fahrenberg, J., Lauer, C. J., Holsboer, F., & Heuser, I. (1995). Cortisol, ACTH, and cardiovascular response to a cognitive challenge paradigm in aging and depression. *The American Journal of Physiology*, *268*(4 Pt 2), R865-873. https://doi.org/10.1152/ajpregu.1995.268.4.R865

Gunnar, M. R. (1987). Psychobiological Studies of Stress and Coping: An Introduction. *Child Development*, *58*(6), 1403–1407. https://doi.org/10.2307/1130680

Gunnar, M. R., Reid, B. M., Donzella, B., Miller, Z. R., Gardow, S., Tsakonas, N. C., Thomas, K. M., DeJoseph, M., & Bendezú, J. J. (2021). Validation of an online version of the Trier Social Stress Test in a study of adolescents. *Psychoneuroendocrinology*, *125*, 105111. https://doi.org/10.1016/j.psyneuen.2020.105111

Hargittai, E., Piper, A. M., & Morris, M. R. (2019). From internet access to internet skills: Digital inequality among older adults. *Universal Access in the Information Society*, *18*(4), 881–890. https://doi.org/10.1007/s10209-018-0617-5

Harvie, H. M. K., Jain, B., Nelson, B. W., Knight, E. L., Roos, L. E., & Giuliano, R. J. (2021). Induction of acute stress through an internet-delivered Trier Social Stress Test as assessed by photoplethysmography on a smartphone. *Stress (Amsterdam, Netherlands)*, *24*(6), 1023–1032. https://doi.org/10.1080/10253890.2021.1995714

Helminen, E. C., Morton, M. L., Wang, Q., & Felver, J. C. (2021). Stress Reactivity to the Trier Social Stress Test in Traditional and Virtual Environments: A Meta-Analytic Comparison. *Psychosomatic Medicine*, *83*(3), 200–211. https://doi.org/10.1097/PSY.0000000000000918

Herten, N., Otto, T., Adolph, D., Pause, B. M., Kumsta, R., & Wolf, O. T. (2016). Enhanced startle responsivity 24 hours after acute stress exposure. *Behavioral Neuroscience*, *130*(5), 521–530. https://doi.org/10.1037/bne0000156

Herten, N., Otto, T., & Wolf, O. T. (2017). The role of eye fixation in memory enhancement under stress—An eye tracking study. *Neurobiology of Learning and Memory*, *140*, 134–144. https://doi.org/10.1016/j.nlm.2017.02.016

Herten, N., Pomrehn, D., & Wolf, O. T. (2017). Memory for objects and startle responsivity in the immediate aftermath of exposure to the Trier Social Stress Test. *Behavioural Brain Research*, *326*, 272–280. https://doi.org/10.1016/j.bbr.2017.03.002

Huneke, N., Rowlatt, H., Hyde, J., McEwan, A., Maryan, L., Baldwin, D., & Garner, M. (2021). *A Modified Trier Social Stress Test to Investigate Social Anxiety using Videoconferencing Software: A Proof-of-Concept study*. PsyArXiv. https://doi.org/10.31234/osf.io/q5rbj

Iuchi, K., Mitsuhashi, R., Goto, T., Matsubara, A., Hirayama, T., Hashizume, H., & Tsumura, N. (2020). Stress levels estimation from facial video based on non-contact measurement of pulse wave. *ARTIFICIAL LIFE AND ROBOTICS*, *25*(3), 335–342. https://doi.org/10.1007/s10015-020-00624-4

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*(9), 1397–1403. https://doi.org/10.1016/j.psyneuen.2010.04.003

Kirschbaum, C. (2021). Why we need an online version of the Trier Social Stress Test. *Psychoneuroendocrinology*, *125*, 105129. https://doi.org/10.1016/j.psyneuen.2021.105129

Kirschbaum, C., Diedrich, O., Gehrke, J., Wüst, S., & Hellhammer, D. (1992). Cortisol and Behavior: The “Trier Mental Challenge Test” (TMCT) — First Evaluation of a New Psychological Stress Test. In A. Ehlers, W. Fiegenbaum, I. Florin, & J. Margraf (Eds.), *Perspectives and Promises of Clinical Psychology* (pp. 67–78). Springer US. https://doi.org/10.1007/978-1-4899-3674-5_7

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

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies” (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6), 115:1-115:35. https://doi.org/10.1145/3457607

Meier, M., Haub, K., Schramm, M.-L., Hamma, M., Bentele, U. U., Dimitroff, S. J., Gärtner, R., Denk, B. F., Benz, A. B. E., Unternaehrer, E., & Pruessner, J. C. (2022). Validation of an online version of the trier social stress test in adult men and women. *Psychoneuroendocrinology*, *142*, 105818. https://doi.org/10.1016/j.psyneuen.2022.105818

Minkley, N., Schröder, T. P., Wolf, O. T., & Kirchner, W. H. (2014). The socially evaluated cold-pressor test (SECPT) for groups: Effects of repeated administration of a combined physiological and psychological stressor. *Psychoneuroendocrinology*, *45*, 119–127. https://doi.org/10.1016/j.psyneuen.2014.03.022

Mitsuhashi, R., Iuchi, K., Goto, T., Matsubara, A., Hirayama, T., Hashizume, H., & Tsumura, N. (2019). Video-Based Stress Level Measurement Using Imaging Photoplethysmography. *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 90–95. https://doi.org/10.1109/ICMEW.2019.0-105

Montero-López, E., Santos-Ruiz, A., García-Ríos, M. C., Rodríguez-Blázquez, R., Pérez-García, M., & Peralta-Ramírez, M. I. (2016). A virtual reality approach to the Trier Social Stress Test: Contrasting two distinct protocols. *Behavior Research Methods*, *48*(1), 223–232. https://doi.org/10.3758/s13428-015-0565-4

Neupert, S. D., Miller, L. M. S., & Lachman, M. E. (2006). Physiological reactivity to cognitive stressors: Variations by age and socioeconomic status. *International Journal of Aging & Human Development*, *62*(3), 221–235. https://doi.org/10.2190/17DU-21AA-5HUK-7UFG

Norden, M., Hofmann, A. G., Meier, M., Balzer, F., Wolf, O. T., Böttinger, E., & Drimalla, H. (2022). Inducing and Recording Acute Stress Responses on a Large Scale With the Digital Stress Test (DST): Development and Evaluation Study. *Journal of Medical Internet Research*, *24*(7), e32280. https://doi.org/10.2196/32280

Norden, M., Mayer, A., Molkenthin, D., Stubenvoll, C., & Drimalla, H. (2022). *Validation of a Digital Stress Test (DST) for the induction of acute stress responses*. https://doi.org/10.17605/OSF.IO/XWU9A

Norden, M., Wolf, O. T., Lehmann, L., Langer, K., Lippert, C., & Drimalla, H. (2022). Automatic Detection of Subjective, Annotated and Physiological Stress Responses from Video Data. *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–8. https://doi.org/10.1109/ACII55700.2022.9953894

Pfeifer, L. S., Heyers, K., Ocklenburg, S., & Wolf, O. T. (2021). Stress research during the COVID-19 pandemic and beyond. *Neuroscience and Biobehavioral Reviews*, *131*, 581–596. https://doi.org/10.1016/j.neubiorev.2021.09.045

Reed, S. C., Oliva, A., Gomez, S. G., & Evans, S. M. (2023). A novel remote TSST procedure reliably increases stress reactivity in cannabis users: A pilot study. *Experimental and Clinical Psychopharmacology*, *31*(1), 7–13. https://doi.org/10.1037/pha0000528

Rimmele, U., Zellweger, B. C., Marti, B., Seiler, R., Mohiyeddini, C., Ehlert, U., & Heinrichs, M. (2007). Trained men show lower cortisol, heart rate and psychological responses to psychosocial stress compared with untrained men. *Psychoneuroendocrinology*, *32*(6), 627–635. https://doi.org/10.1016/j.psyneuen.2007.04.005

Roldán-Rojo, T. A., Rendón-Veléz, E., & Carrizosa, S. (2021). Stressors and Algorithms Used for Stress Detection: A Review. *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–8. https://doi.org/10.1109/ACII52823.2021.9597456

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. *Psychosomatic Medicine*, *70*(7), 787–796. https://doi.org/10.1097/PSY.0b013e3181810658

Shiban, Y., Diemer, J., Brandl, S., Zack, R., Mühlberger, A., & Wüst, S. (2016). Trier Social Stress Test in vivo and in virtual reality: Dissociation of response domains.

International Journal of Psychophysiology, *110*, 47–55. https://doi.org/10.1016/j.ijpsycho.2016.10.008

Shore, A. (2022). Talking about facial recognition technology: How framing and context influence privacy concerns and support for prohibitive policy. *Telematics and Informatics*, *70*, 101815. https://doi.org/10.1016/j.tele.2022.101815

Tonguç, G., & Ozaydın Ozkara, B. (2020). Automatic recognition of student emotions from facial expressions during a lecture. *Computers & Education*, *148*, 103797. https://doi.org/10.1016/j.compedu.2019.103797

Turner-Cobb, J. M., Asif, M., Turner, J. E., Bevan, C., & Fraser, D. S. (2019). Use of a non-human robot audience to induce stress reactivity in human participants. *Computers in Human Behavior*, *99*, 76–85. https://doi.org/10.1016/j.chb.2019.05.019

Wang, W., Tran, D., & Feiszli, M. (2020). What Makes Training Multi-Modal Classification Networks Hard? *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12692–12702. https://doi.org/10.1109/CVPR42600.2020.01271

Westenberg, P. M., Bokhorst, C. L., Miers, A. C., Sumter, S. R., Kallen, V. L., van Pelt, J., & Blöte, A. W. (2009). A prepared speech in front of a pre-recorded audience: Subjective, physiological, and neuroendocrine responses to the Leiden Public Speaking Task. *Biological Psychology*, *82*(2), 116–124. https://doi.org/10.1016/j.biopsycho.2009.06.005

Wolf, O. T., Schulte, J. M., Drimalla, H., Hamacher-Dang, T. C., Knoch, D., & Dziobek, I. (2015). Enhanced emotional empathy after psychosocial stress in young healthy men. *Stress*, *18*(6), 631–637. https://doi.org/10.3109/10253890.2015.1078787

Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., & Sahebkar, A. (2017). The impact of stress on body function: A review. *EXCLI Journal*, *16*, 1057–1072. https://doi.org/10.17179/excli2017-480

Zhang, H., Feng, L., Li, N., Jin, Z., & Cao, L. (2020). Video-Based Stress Detection through Deep Learning. *Sensors (Basel, Switzerland)*, *20*(19), 5552. https://doi.org/10.3390/s20195552

Eidesstattliche Versicherung

„Ich, Matthias Norden, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: **„A Novel Approach to Inducing and Measuring Acute Stress Responses: Development, Evaluation and Application Potential of a Digital Stress Test“** *(„Ein innovativer Ansatz für die Induktion und Messung von Akuten Stressreaktionen: Entwicklung, Evaluierung und Anwendungspotential eines Digitalen Stress Tests“)* selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.
Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.og) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum Unterschrift

Anteilserklärung an der erfolgten Publikation

Matthias Norden hatte folgenden Anteil an der folgenden Publikation, die der medizinischen Promotion zu Grunde liegt:

> **Matthias Norden**, Amin Gerard Hofmann, Martin Meier, Felix Balzer, Oliver T Wolf, Erwin Böttinger, Hanna Drimalla, „Inducing and Recording Acute Stress Responses on a Large Scale With the Digital Stress Test (DST): Development and Evaluation Study" , Journal of Medical Internet Research, 2022

Journal Impact Factor (2021): 7.076
Beitrag im Einzelnen:

- Literaturrecherche zur akuten Stressreaktion inkl. physiologischer Grundlagen, Folgen und klassischen psychologischen Forschungsmethoden
- Konzeptualisierung des Digitalen Stress Tests (DST) und des Control – DST (C-DST) anhand der in der Literatur beschriebenen Stressinduktionsprinzipien
- Mitarbeit an der technischen Implementierung der Prototypen von DST und C-DST (JavaScript Programmierung, Front-End Planung, Integration in Study Management System JATOS)
- Planung der Online-Studie zur Evaluation des Stressinduktions-Potentials Mitgestaltung des Ethikantrags für die Durchführung der Studie
- Vorbereitung der Fragebögen auf der Fragebogen-Plattform SoSci-Survey - Rekrutierung der Teilnehmenden für die Online-Studie
- Statistische Auswertung der Fragebögen-Daten und Stressinduktionsdaten (Ergebnis-Teil Publikation)
- Schriftliche Ausarbeitung der Publikation mit Korrekturen durch Ko-Autor*innen
- Erstellung aller Tabellen und Figuren der Publikation; Figure 2 der Publikation nach grobem Entwurf durch Projektmitarbeitenden
- Durchführung des Publikationsprozesses inkl. Beantwortung und Korrekturen im Peer-Review-Verfahren mit Absprache der KoautorInnen

____

Unterschrift, Datum und Stempel des/der erstbetreuenden Hochschullehrers/in

____

Unterschrift des Doktoranden/der Doktorandin

Druckexemplar der Publikation

JOURNAL OF MEDICAL INTERNET RESEARCH Norden et al

Original Paper

Inducing and Recording Acute Stress Responses on a Large Scale With the Digital Stress Test (DST): Development and Evaluation Study

Matthias Norden[1,2,3]; Amin Gerard Hofmann[1,2]; Martin Meier[2]; Felix Balzer[3]; Oliver T Wolf[4]; Erwin Böttinger[2,5]; Hanna Drimalla[1,2]

[1]Faculty of Technology, Bielefeld University, Bielefeld, Germany

[2]Digital Health Center, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

[3]Institute of Medical Informatics, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[4]Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany

[5]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York City, NY, United States

Corresponding Author:
Hanna Drimalla
Faculty of Technology
Bielefeld University
Postfach 10 01 31
Bielefeld, 33619
Germany
Phone: 49 521 106 12043
Email: drimalla@techfak.uni-bielefeld.de

Abstract

Background: Valuable insights into the pathophysiology and consequences of acute psychosocial stress have been gained using standardized stress induction experiments. However, most protocols are limited to laboratory settings, are labor-intensive, and cannot be scaled to larger cohorts or transferred to daily life scenarios.

Objective: We aimed to provide a scalable digital tool that enables the standardized induction and recording of acute stress responses in outside-the-laboratory settings without any experimenter contact.

Methods: On the basis of well-described stress protocols, we developed the Digital Stress Test (DST) and evaluated its feasibility and stress induction potential in a large web-based study. A total of 284 participants completed either the DST (n=103; 52/103, 50.5% women; mean age 31.34, SD 9.48 years) or an adapted control version (n=181; 96/181, 53% women; mean age 31.51, SD 11.18 years) with their smartphones via a web application. We compared their affective responses using the international Positive and Negative Affect Schedule Short Form before and after stress induction. In addition, we assessed the participants' stress-related feelings indicated in visual analogue scales before, during, and after the procedure, and further analyzed the implemented stress-inducing elements. Finally, we compared the DST participants' stress reactivity with the results obtained in a classic stress test paradigm using data previously collected in 4 independent Trier Social Stress Test studies including 122 participants overall.

Results: Participants in the DST manifested significantly higher perceived stress indexes than the Control-DST participants at all measurements after the baseline ($P<.001$). Furthermore, the effect size of the increase in DST participants' negative affect ($d=0.427$) lay within the range of effect sizes for the increase in negative affect in the previously conducted Trier Social Stress Test experiments (0.281-1.015).

Conclusions: We present evidence that a digital stress paradigm administered by smartphone can be used for standardized stress induction and multimodal data collection on a large scale. Further development of the DST prototype and a subsequent validation study including physiological markers are outlined.

(J Med Internet Res 2022;24(7):e32280) doi: 10.2196/32280

KEYWORDS

stress induction; smartphone; stress reactivity; Trier Social Stress Test; TSST; remote; video recording; acute stress; digital health; mobile health; mHealth; mobile phone

Introduction

Relevance and Rationale

Psychosocial stress is a major risk factor for the development of physical and mental illnesses, including hypertension, depression, and anxiety [1]. Valuable insights into its causes and consequences have been gained through experimental stress paradigms during which acute stressors are used to induce a psychosocial stress reaction. For example, such stress induction paradigms have been successfully used to investigate the effects of acute stress on the brain [2], hormonal and inflammatory reactivity [3], memory [4], and social cognition and behavior [5].

Applying controlled stress induction paradigms also enables the investigation of prevention and intervention strategies. For example, in a recent study, Het et al [6] used a classic stress paradigm to study the effects of an inpatient treatment on acute stress reactivity in women with eating disorders. In addition, controlled stress induction procedures play an important role in the development of objective stress detection methods [7,8] as they strongly rely on highly qualitative and representative data sets obtained through stress induction experiments [9].

Current Stress Paradigms and Their Limitations

Currently, most stress induction paradigms are limited in their scalability (ie, applicable across a large number of participants and distances) and, thus, cannot be easily used to gather large volumes of stress-related data. Furthermore, many of these have not been replicated outside the laboratory to verify the laboratory findings in outside-the-laboratory settings [10]. To overcome these limitations, a new standardized and validated stress induction paradigm is needed.

The Trier Social Stress Test (TSST) [11] is considered the gold standard in human experimental stress research, having been applied >4000 times including different populations and age groups [12]. Participants have to complete a 5-minute mock job interview and a 5-minute mental arithmetic task in front of an evaluating committee. This procedure requires a laboratory setup, an experimenter, and 2 actors playing the committee, making the TSST costly and unfeasible for large-scale application. In addition, the impact of the different methodological elements (eg, panel composition) on the stress reaction and the relatively small sample sizes complicate the reproducibility of the findings [13-15]. Furthermore, the experimental setting might lead to stress responses that differ from acute stress experienced in daily life.

Several adaptations have been made to provide less costly and laborious versions, but they still require human resources (eg, TSST for groups) or additional equipment (eg, virtual reality TSST or e-TSST) and have not been tested in nonlaboratory settings. Recently, 2 studies applied a web-based version of the TSST during which adolescent [16] or adult [17] participants joined judges and experimenters on a web-based videoconferencing platform without any in-person assessment. The responses to these web-delivered versions were consistent with standard in-person responses although the paradigm was conducted remotely. This highlights the possibility of assessing stress reactivity outside a research laboratory. However, the entire procedure still depends on live interactions between the committee, the participant, and the experimenter.

Stressors that enable the investigation of stress responses without direct experimenter contact have been developed for imaging scenarios [18]. Using their Imaging Paradigm for Evaluative Social Stress, Fehlner et al [19] showed that delivering short spoken answers to selected topics in front of a prerecorded audience and additional framings induced robust stress responses. This indicates that psychosocial stress can also be induced by making the participants believe they are exposed to some kind of social evaluation without direct experimenter interaction.

The Montreal Imaging Stress Task (MIST) [20] supports this assumption. It comprises computerized mental arithmetic tasks with an induced failure component and social pressure elements. However, these paradigms have only been tested within imaging laboratory settings where experimenters were still present and performed potentially stressful measurements. Thus, the stress induction might be influenced by the imaging setting and the experimenter's role during the procedure.

Many other well-described stress paradigms (eg, the CO2 challenge test and the socially evaluated cold pressor test [21]) are dependent on laboratory settings, build on physical stressors, and require human resources or additional equipment [22]. Other paradigms (eg, the Paced Auditory Serial Addition Task [23] and Stroop test [24]) would theoretically be applicable outside the laboratory but lack the possibility to collect multimodal behavior data (eg, facial expressions and voice recordings) of the stress response. To the best of our knowledge, there is currently no standardized and validated digital stress paradigm that can be carried out without an experimenter and collect multimodal video data of participants in stressed conditions. Therefore, we conceptualized and developed a completely digital stress test to address the need for an innovative, standardized, and validated stress induction protocol.

Digital Stress Test

The Digital Stress Test (DST) is primarily intended as a digital research tool. Importantly, we did not aim to develop a direct stress measurement or therapeutic tool. Instead, the DST enables researchers to gain additional insights into acute stress responses by making stress studies scalable and transferable to outside-the-laboratory settings and collecting stress-relevant video data at the same time. Thus, the DST is designed as an easy-to-use smartphone web application where participants conduct the study (via the internet) without any direct communication with researchers or additional resources required (ie, wearables or native app downloads).

It combines different well-known stress induction principles of classic stress paradigms and adapts them to a digital setting. According to a meta-analysis of psychological stress paradigms by Dickerson and Kemeny [25], a robust and reliable stress response can be induced by acute or chronic threats to social status, particularly when conditions are uncontrollable. Most likely, this would occur when failure or poor performance could reveal a lack of ability. Both principles have been proven effective in state-of-the-art stress paradigms and will be used as the basis for the digital stress induction paradigm.

Second, the DST aims to collect multimodal behavior data (ie, facial and voice cues) that can be used to build a basic data set for further (machine learning) analysis. Therefore, the embedded stress induction procedure will include a naturalistic speaking part (ie, comparable with daily speaking).

Objectives and Hypotheses

The aim of this study was to develop the first prototype of a DST web application and assess its feasibility as well as its stress induction potential. Therefore, we also provided a neutral version called the Control-DST (C-DST) that can be used similarly in web-based settings. We hypothesized that the DST would elicit a stronger stress response compared with the neutral condition. In addition, we placed our results in the context of previous studies conducting the gold-standard paradigm (TSST).

This paper is organized as follows: in the *Methods* section, we describe the concept and development of the DST and its control version. Furthermore, we provide details of the large-scale web-based study conducted to evaluate the feasibility and stress induction potential of the DST. In the *Results* section, we present statistical evidence for the stress induction potential of the DST. Finally, in the *Discussion* section, we discuss our results and potential limitations in light of previous work and outline plans for future research.

Methods

Concept and Development of the DST

We first describe the underlying stress induction paradigm as well as its adaptation for the development of a suitable control condition. We include illustrations of the first DST and C-DST prototypes and outline the technological aspects. Before starting the web-based evaluation study, we conducted a pilot study to finalize the prototypes based on participants' feedback.

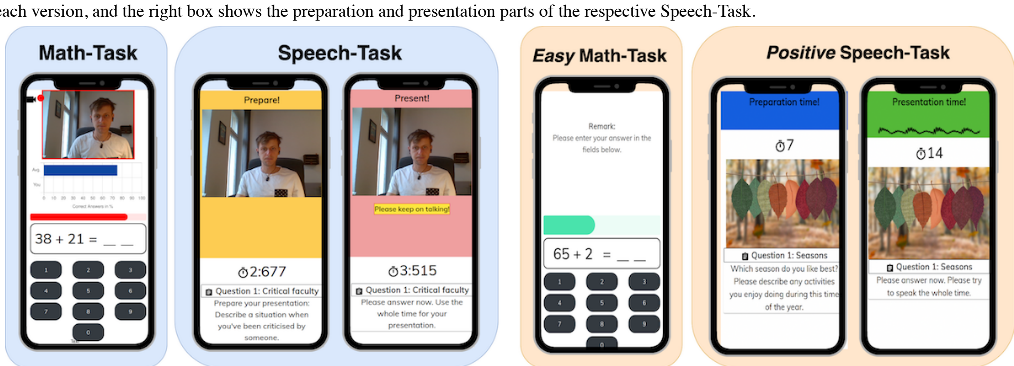
Concept of the DST

Overview

The paradigm consists of an arithmetic calculation and a free speech part and is framed as a cognitive-verbal performance test. Screenshots of the DST and its control condition are shown in Figure 1. The complete web application procedures can be seen in Multimedia Appendices 1 and 2. Presentation versions of the most recent DST and C-DST without any data saving can be found at their respective websites [26,27].

To elicit a robust acute psychosocial stress reaction, the DST procedure comprises multiple elements of social-evaluative threat and uncontrollability [11,25].

Figure 1. Screenshots of the Digital Stress Test (blue boxes) and Control - Digital Stress Test (orange boxes) tasks. The left box illustrates the Math-Task in each version, and the right box shows the preparation and presentation parts of the respective Speech-Task.



Framing

The DST is introduced as a research tool for "behavior analysis while performing a cognitive-verbal performance test," indicating that the individual performance of the participant is tested. To further increase the social-evaluative threat, they are informed that they will be recorded through the front camera of their smartphones and that these recordings are being analyzed to assess their individual resilience.

The participants record a short test video that claims to calibrate the implemented algorithm and shall increase the credibility of the automated behavior analysis.

The cognitive task is framed as a simple calculation task that a fake comparison group (based on age and gender) apparently solved with an average of 75% correct answers. This intends to emphasize the expected results and introduce the participants to the permanent social comparison in the upcoming calculation task, as done in the MIST [20].

Arithmetic Calculation Task (*Math-Task*)

The task comprises elements of the MIST [20] protocol adapted to the smartphone setting and enhanced with several other stress-inducing elements. After a countdown, the participants are required to solve simple calculation tasks consisting of addition, subtraction, multiplication, and division of 2 numbers ranging from 1 to 99 with solutions ranging from 1 to 99. The participants need to type their solution on a number field within the given time limit. If the response is wrong or no response is recorded within the time limit, negative feedback is presented ("Wrong answer!" or "Too slow!") and the background color changes to red. After a correct response, the next calculation task is presented immediately. The time limit for each calculation is marked using a red expiring progress bar.

A continuous failure rate is being provoked. For the first task, the time limit is set to 3 seconds. If the participant answers a series of 3 consecutive arithmetic tasks correctly, the time limit is shortened by 10%. In addition, for the following 4 tasks, the numbers of the input field are swapped randomly to increase the difficulty and uncontrollability. If the participant answers a series of 3 consecutive tasks incorrectly (or not at all), the time limit is extended by 10%. If the participant does not give any input for 5 consecutive tasks, feedback indicating the relevance of the study is displayed, and the next task is chosen to be easily solvable (ie, a summation task). This intends to ensure ongoing participation.

During the Math-Task, the percentage of correct responses is continuously displayed and compared with the fixed average of the participants' age- and gender-related groups in a bar chart. As the achieved percentage of correct answers in the comparison group was claimed to be 75%, which usually exceeds the current percentage of the participant because of the implemented difficulty, this continually reminds the participant of failing.

In addition, the front camera of the smartphone is activated, and the recorded video is displayed directly on the upper half of the screen during the entire Math-Task. This intends to remind the participants that they are being recorded and apparently analyzed while failing in a school-like performance task.

The participants do not know how long the Math-Task takes to increase a feeling of uncontrollability. After 1.5 minutes, the Math-Task automatically stops. The participants see their final percentage of correct answers compared with the fabricated age- and gender-related average and are reminded that "only serious results can be used for this study," emphasizing the relevance of the participants' performance.

Free Speech Task (*Speech-Task*)

The second part is the Speech-Task, which further extends the social-evaluative threat through a presentation-like situation and enables the recording of stress-relevant voice cues. The participants are reminded that their verbal skills will be assessed. They are instructed to prepare structured and convincing verbal answers to standard job interview questions. They are not told how many questions will follow, making the length of this task unpredictable.

The Speech-Task includes 3 inconvenient answering scenarios (eg, "Describe a situation when you've been criticized by someone!") that are based on a previous study by Fehlner et al [19]. For each scenario, they are given 10 seconds to prepare and 20 seconds to present their speech. The participants are reminded to use the entire time for their presentation.

A countdown indicates the time for preparation and presentation, intending to pressure the participants. During their presentations, the background color of the entire screen blinks red to visually distract and agitate the participants.

The smartphone's front camera is activated, and the recorded video is displayed on the upper half of the screen during the preparation and presentation periods. In addition, a voice visualization is included in the presentation parts. After 1 second without recorded noise input signal, the participant is reminded to keep on talking, increasing the credibility of the behavior analysis and pressuring the participants, as done by the experimenters in the TSST paradigm [11,13].

Concept of the C-DST

Overview

We also developed a control version of the DST that resembles its structure and procedure but differs in terms of the stress induction elements (Figure 1, right side). We changed the tasks and framings to be less stressful, as done for the placebo TSST [28] and friendly TSST [29]. The provided information on the study's background, privacy, and data protection aspects, as well as the performance task framing in the beginning, remains exactly the same to have a comparable baseline. The differences are outlined in the following sections.

Friendly Framing

The participants are informed that they are part of a control group and that they will not be video recorded. No recording of a test video or any additional framing of an automated behavior analysis takes place. The participants are not told that their individual performance results will be compared with those of other participants, and no fictive average result scores are displayed.

Easy Math-Task

The calculation tasks in the C-DST are generated in the same way as in the DST but only with summation tasks. The time limit for the first task is set to 5 seconds. The time adaptation algorithm is designed to enable more correct responses—as soon as the participant answers 1 task incorrectly (or not at all), the time limit is extended by 10%.

Only if the participant answers a series of 4 consecutive tasks correctly the time limit is shortened by 10%. In contrast to the DST, the numbers of the input field are not swapped for the following task.

The provided feedback is chosen to be encouraging (ie, the screen color changes to green for correct answers and does not change for wrong answers). The time limit is marked using a green progress bar. Neither a fake comparison with other participants' results nor any live recording through the front camera is displayed.

Positive Speech-Task

As opposed to the DST Speech-Task, the *positive* Speech-Task is not introduced as an assessment of the participants' verbal skills. The answering scenarios include only neutral topics (eg, "Which season do you like best? Please describe any activities you enjoy doing during this time of the year!").

Instead of displaying the live-recorded video of the front camera on the upper half of the screen, a neutral image suiting the question is shown. No further distraction through a red blinking background takes place, and the colors are chosen to be calming.

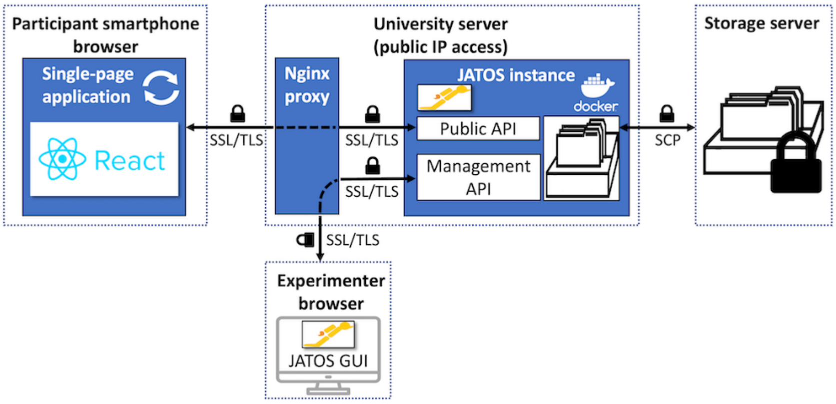
Technological Aspects

The system architecture of the applications is shown in Figure 2. The DST and C-DST were developed as single-page web applications using the JavaScript framework React.js. The source code of the most recent versions is publicly accessible at the website [30]. The applications run on standard browsers and are hosted on a university server that allows for public IP access using the open-source study management system JATOS [31] within a Docker container. JATOS exposes a public application programming interface (API) that is called with a wrapper library and handles requests from a participant's browser (eg, fetch and upload data). In addition, it provides a management API to handle requests from the experimenter's browser via the JATOS graphical user interface. More detailed information on the JATOS architecture can be found in the study by Lange et al [31].

In this study, only fully anonymized data were collected. We disabled the recording of videos but only streamed them within the participant's smartphone browser as the focus of this study was testing and validating the digital stress induction procedure. Owing to the capability and future plans to also collect sensitive and potentially identifiable video data, we implemented several security measures.

Nginx (Nginx, Inc) is used on the publicly reachable university server to ensure Secure Sockets Layer encryption, and it only responds to https requests for calls to both the public and management JATOS APIs. Participant data are only temporarily stored on the web server and directly transferred to a secure storage server via secure copy protocol after the test ends. All (remaining) data are deleted automatically from the web server in short time intervals. We have already received ethics approval for our data storage concept. For the future, we also plan to implement a client-side encryption of participant data files that takes place already within the web applications and can only be decrypted using private keys from the secure storage server.

Figure 2. System architecture of the test applications. The Digital Stress Test and Control - Digital Stress Test work as single-page web applications within the participant's smartphone browser (left side). The single-page applications are hosted on an web-based reachable university server (center) using the open-source web study management system JATOS [28] within a Docker container. API: application programming interface; GUI: graphical user interface; SCP: secure copy protocol; SSL: secure socket layer; TLS: transport layer security.



Pilot Study

We conducted a pilot study with 49 participants performing either the DST (21/49, 43%) or the C-DST (28/49, 57%) web application. On the basis of their feedback, we adjusted major usability issues that were caused by different browsers and smartphones and fixed technological bugs. We aimed for a comprehensive study introduction and consenting and debriefing information and modified the wording accordingly.

Evaluation of the DST

To assess the feasibility and stress induction potential of the DST, we first conducted a large web-based study. Participants in this web-based study performed either the DST or the C-DST and filled out several questionnaires regarding their affective responses. The effect sizes of the affective changes indicated by the DST participants in this web-based study were then compared with results obtained in previous studies performing the laboratory gold-standard paradigm (TSST).

Participants and Recruitment for the Web-Based Study

Overview

Participants were recruited via web-based publication of the study link in the university and study participation mailing lists, social networks (eg, Twitter and Facebook), podcasts, and websites. The study was conducted for 2 weeks, from February 10 to February 24, 2021. Within this period, 547 participants performed either the DST (300/547, 54.8%) or C-DST (247/547, 45.2%). For the evaluation of subjective stress parameters, we excluded participants with incomplete tests (229/547, 41.9%), previous self-reported knowledge of the framing (13/547, 2.4%), self-reported usability issues (19/547, 3.5%), or unrealistic procedure duration (2/547, 0.4%).

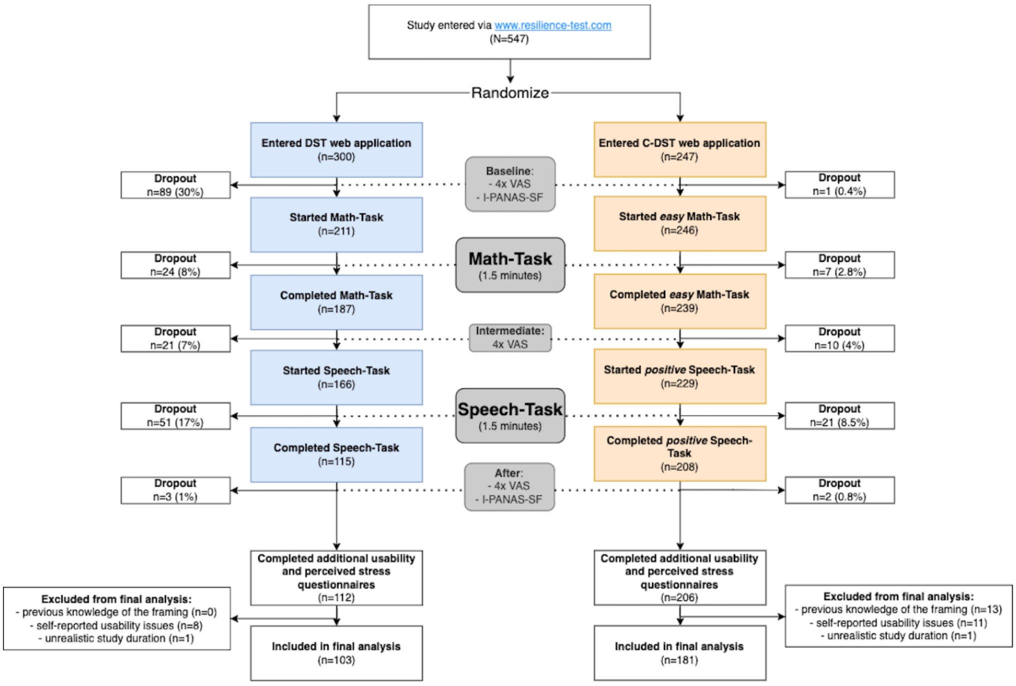
Web-Based Study Procedure

The design of the web-based study is shown in Figure 3. The entire procedure takes place on the screen of the participants' smartphones and takes approximately 5 to 10 minutes. Using the provided study link, the participants were randomly forwarded to either the stress or control paradigm. We adjusted the randomization algorithm to prefer the DST when we analyzed the dropout rates after the first week. However, most participants (479/547, 87.6%) performed the study within the first week because of the publication in a widespread German politics podcast.

During the pretest part of the web application, the participants were introduced to the study background, upcoming procedure, and privacy and data protection aspects. Assessments of the perceived stress level took place before, between, and after the 2 tasks using built-in questionnaires.

After completing both tasks, the participants were debriefed and linked to additional usability and follow-up questionnaires on an external website [32]. The participants could quit the study at any time (eg, by closing the browser).

Figure 3. Web-based study design, drop-outs, and collected data. The participants are randomly assigned to either the Digital Stress Test (blue boxes) or Control - Digital Stress Test (orange boxes) web application using the provided study link. The participants answer the same questionnaires within the respective web application and additional questionnaires on an external website afterward. I-PANAS-SF: international Positive and Negative Affect Schedule Short Form; VAS: Visual Analogue Scale.



Web-Based Study Data Collection

General demographic information, including age and gender, was obtained in the first part of the session. In addition, previous participation was asked about to exclude duplicated results with previous knowledge and confounding factors.

To assess the perceived stress levels, the participants completed several self-report questionnaires (Textbox 1).

To further compare the feasibility of the 2 paradigms, several pieces of metadata were stored during the procedure. This included the performance during the Math-Task and the study progress (ie, how much time the participants spent on different parts of the application and at which part they cancelled).

The participants were linked to additional questionnaires at the end of the web application. These included several questions on usability aspects (eg, problems with the correct camera and problems with the Visual Analogue Scale [VAS]) as well as the opportunity to provide open feedback. Furthermore, we asked the participants to rate their perceived stress regarding specific parts of the applications on a VAS.

Textbox 1. In-app self-report questionnaires completed by the participants.

Self-report questionnaires

- The international short form of the Positive and Negative Affect Schedule (PANAS) [33] was applied to assess positive and negative affect in the beginning (baseline) and after solving both tasks (posttest assessment). The PANAS is a well-validated and reliable tool to assess the participants' mood that has been applied in various studies on mood changes [34]. The participants indicate the intensity of 10 feelings and emotions on a 5-point Likert scale. The items can be subdivided into negative affect (NA; consisting of 5 items) and positive affect (PA; consisting of 5 items). We used the mean scores for both affects and normalized them for the number of items (ie, PA and NA outcomes ranging from 1 to 5 for each time point).
- Visual Analogue Scales (VAS) regarding 4 different dimensions of stress (feeling *stressed*, *frustrated*, *overstrained*, and *ashamed*) were obtained in the beginning, between the 2 tasks, and in the end. The VAS is a common instrument to measure characteristics that cannot be easily measured directly and is often used for pain, stress, or other subjective experiences [35]. The participants indicate how much they are perceiving specific feelings at the current moment by choosing a point on a fixed-size horizontal line where the ends are defined as the extremes (eg, *not at all* and *very much*). The VAS score is then determined by measuring the relative distance from the left end of the line to the participants' chosen point.

External Evaluation With the TSST

Data from 122 participants who underwent the traditional TSST procedure were previously collected at Ruhr University Bochum in 4 independent studies [34,36-38]. The procedures included assessments of the affective responses using the Positive and Negative Affect Schedule (PANAS). We used the archived data and compared the effects of the TSST on the participants' affective responses with the responses indicated by the DST participants in this study.

Statistical Analysis

Statistical analysis was performed using Python 3.7 (Python Software Foundation) with the pandas, statsmodels, and pingouin libraries. We assessed the distributions for normality and homogeneity of variances using Shapiro-Wilk and Levene tests, respectively. The participants' affective responses were analyzed using mixed-model ANOVAs for repeated measurements with the factor *time* (baseline and after for the PANAS; baseline, intermediate, and after for the VAS) and the between-subject factor *group* (DST group vs C-DST group) separately for the PANAS and VAS scales. Owing to their robustness against deviations from the normality assumption [39], we also used ANOVAs for nonnormally distributed data. Greenhouse-Geisser corrections for df were applied where sphericity could not be assumed. Post hoc tests were performed using Bonferroni-adjusted Welch *t* test for different sample sizes and nonhomogeneity of variances [40].

To further analyze the DST parts regarding their stress induction potential, we calculated the mean VAS scores for every part of the DST or C-DST evaluated in the posttest questionnaire and descriptively ranked them.

To compare the affective responses of the participants performing the DST with those of the participants who underwent the TSST in previous studies, we analyzed the normalized scores for the PANAS positive and negative affect subscales using a 2-step meta-analysis. Therefore, we first performed paired *t* tests on the normalized pre- and post-PANAS scores for each of the TSST studies separately and calculated standardized effect sizes. Afterward, we computed a combined effect size for all TSST studies by assigning weights based on the inverse of the change score variance to the individual effect sizes of the respective studies [41] and compared it with the standardized effect size observed in DST participants.

In all analyses reported, we used 2-tailed comparisons with a P value of <.05 as the significance criterion. The effect size was reported using partial η^2 for ANOVA and Cohen d_z for paired t tests [42].

Ethics Approval

Ethics approval for the study was granted by the University of Potsdam (application 33/2020), and the study was conducted in accordance with the General Data Protection Regulation. As this web-based study was conducted without experimenter supervision, special care was taken to ensure General Data Protection Regulation– and ethics-compliant informed consent, debriefing, and study cancellation process.

Results

Participants and Dropouts

Overall, 103 individuals completed the DST (50/103, 48.5% men; 52/103, 50.5% women; and 1/103, 1% other; mean age 31.34, SD 9.48 years), and 181 individuals completed the C-DST (83/181, 45.9% men, 96/181, 53% women, and 2/181, 1.1% other; mean age 31.51, SD 11.18 years). Most participants had a high level of education in both the DST (65/102, 63.7% had a university degree and 33/102, 32.4% had a high school degree) and C-DST groups (112/175, 64% had a university degree and 52/175, 29.7% had a high school degree). More details on the study participants' ages and educational backgrounds can be found in Figures S1 and S2 in Multimedia Appendix 3.

The average time taken to complete the procedure was 7.69 (SD 1.35) minutes for the DST and 6.53 (SD 1.05) minutes for the C-DST. Most participants (263/284, 92.6%) did not report any usability issues.

Beyond the completed studies, 247 individuals started the study but dropped out. For the C-DST, 83.4% (206/247) of the initial participants completed the procedure, whereas 37.3% (112/300) completed the DST paradigm. The dropout rates at different time points during the procedure are shown in Figure 3. Most DST participants who did not finish the study had already dropped out before starting the Math-Task. Participants who did not complete the study were not included in the following analyses.

DST Versus C-DST

The DST and C-DST participants' affective responses indicated in the PANAS questionnaires are shown in Figure 4. We found a significant main effect for the factor *group* ($F_{1,282}$=5.83; P=.02; η_p^2=0.02) accompanied by a significant *group×time* interaction effect ($F_{1,282}$=31.37; P<.001; η^2=0.10) in the PANAS negative affect subscale. Post hoc analyses for the *group* effect showed that the participants' overall reported negative affect was higher in the DST (mean 1.70, SD 0.55) than in the C-DST (mean 1.54, SD 0.57) group (P<.001). Post hoc tests for the *group×time* interaction effect revealed that the participants' indicated negative affect did not significantly differ between the DST (mean 1.57, SD 0.56) and C-DST (mean 1.58, SD 0.58) groups in the baseline measurements (P=.99) but was significantly higher in the posttest assessments for DST (mean 1.84, SD 0.7) than for C-DST (mean 1.49, SD 0.64) participants (P<.001).

Conducting separate mixed-model ANOVAs for the participants' indicated positive affect, we did not find significant *group* (P=.40) or *group×time* interaction (P=.51) effects, but we did find a significant *time* effect ($F_{1,282}$=0.43; P<.001; η_p^2=0.002). The post hoc analysis showed that, overall, perceived positive affect increased in study participants (baseline: 3.02 –0.65 to +0.65; after the procedure: 3.34 –0.76 to +0.76; P<.001).

The participants' responses to the 4 different VASs are shown in Figure 5. Regarding the *stress* scale, mixed-model ANOVAs revealed significant main effects for the factors *group* ($F_{1,282}$=14.42; P<.001; η_p^2=0.05) and *time* ($F_{2,564}$=75.11; P<.001; η_p^2=0.21) that were moderated by a significant *group×time* interaction effect ($F_{2,564}$=14.28; P<.001; η_p^2=0.05). Post hoc analyses for the *group* effect revealed that overall reported stress responses were higher for the DST (mean 42.39, SD 20.87) than for the C-DST (mean 32.79, SD 20.27) participants (P<.001). Post hoc tests for the *time* effect showed that participants' perceived stress significantly increased over the Math-Task (baseline: 32.92 –25.48 to +25.48; intermediate: 46.76 –26.01 to +26.01; P<.001) and decreased over the Speech-Task (intermediate: 46.76 –26.01 to +26.01; after the procedure: 29.15 –25.96 to +25.96; P<.001). Analyzing the *group×time* interaction, the participants' indicated VAS scores were significantly higher in the DST group than in the C-DST group at all time points after the baseline measurements (P<.001 in all cases). Furthermore, we found very similar patterns for the 3 other stress-related attributes (*frustration*, *shame*, and *overstrain*) conducting separate mixed-model ANOVAs and post hoc tests (Multimedia Appendix 4).

Figure 4. Negative (A) and positive (B) affect indicated in the Positive and Negative Affect Schedule (PANAS) subscales at baseline and posttest assessments for each participant in the Digital Stress Test (blue) and Control - Digital Stress Test (orange). A significant interaction between time and group was found for the negative but not the positive affect subscale. Digital Stress Test participants' negative affect was significantly higher at post-test assessment than Control - Digital Stress Test participants' negative affect (**P<.001 in post hoc Welch t test), whereas baseline scores did not significantly differ.

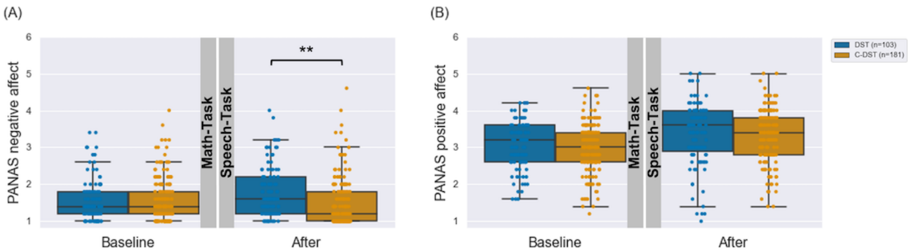
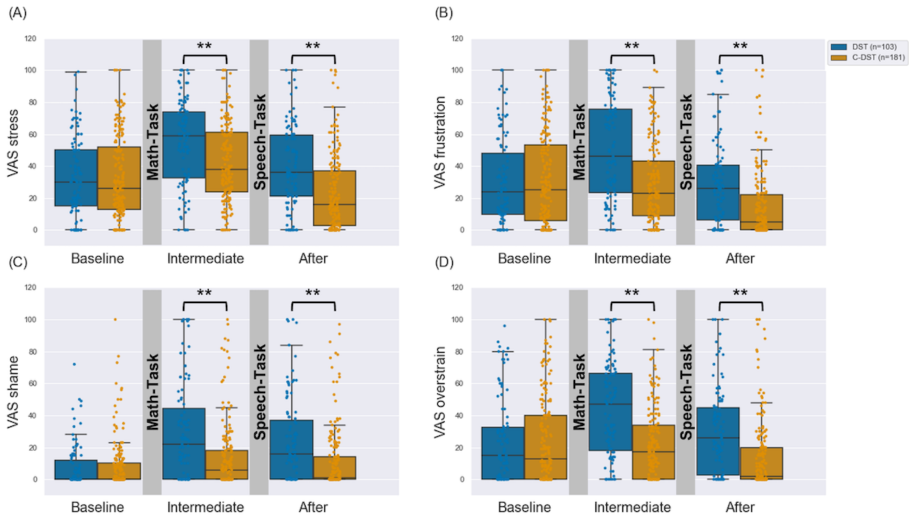


Figure 5. Visual Analogue Scale (VAS) responses for 4 different stress-related affect dimensions (A-D) of the Digital Stress Test (blue) and Control - Digital Stress Test (orange) groups at different times during the procedure. A significant interaction between time and group was found for all VAS scores. Subjective stress indexes were significantly elevated in the Digital Stress Test group compared with the Control Digital Stress test group at all time points after the baseline measurements (**P<.001 in post hoc Welch t test).



Analysis of Stress Elements

The results of the poststudy stress perception questionnaire are summarized in Table 1. According to the participants, the Math-Task was the most stressful element of the DST when compared with the framings in the beginning and with the Speech-Task.

Regarding the 2 tasks, the participants indicated the highest stress perception for the time pressure, whereas the social-evaluative component of being recorded through the front camera was not perceived as that stressful. Regarding the Math-Task, randomly swapping the input field after having correctly solved 3 calculation tasks seemed to induce a high level of perceived stress in the DST participants. Participants of the C-DST also rated the Math-Task and the implemented time limits as the most stressful elements of this version.

Table 1. Different parts of the Digital Stress Test (DST) and Control-Digital Stress Test (C-DST) and perceived stress levels sorted from highest to lowest indicating stress experience for each part of the DST paradigm.

Element in DST and C-DST and subcategory	Perceived stress—DST feedback (VAS[a]), mean (SD)	Perceived stress—C-DST feedback (VAS), mean (SD)
Framing		
Participation in a performance test	65.4 (23.8)	46.7 (25.2)
Behavior analysis through algorithm	40.6 (27.5)	N/A[b]
Math-Task (overall)	77.9 (18.4)	63.6 (23.7)
Time limit	88.8 (13.5)	73.3 (21.7)
Random input field swap	88.6 (15.6)	N/A
Feedback after every calculation task	67.3 (24.0)	40.8 (27.8)
Task difficulty	62.8 (26.3)	42.4 (23.0)
Live comparison with other participants	59.8 (28.5)	N/A
Personal performance	58.0 (29.0)	29.4 (22.3)
Front camera activation	39.7 (28.6)	N/A
Speech-Task (overall)	46.2 (23.3)	25.5 (19.6)
Preparation periods	45.5 (25.8)	19.6 (22.4)
Time limits	45.1 (27.3)	22.6 (24.8)
Questions	43.6 (23.8)	20.2 (22.7)
Front camera activation	37.1 (25.9)	N/A
Audio visualization of the voice	30.2 (24.2)	18.0 (21.8)

[a]VAS: Visual Analogue Scale.

[b]N/A: not applicable.

DST Versus TSST

To further evaluate the stress induction potential of the DST, we performed a 2-step meta-analysis and compared the effects of the DST with findings of 4 previously conducted TSST studies. The results of each study are shown in Table 2. The sample sizes of the TSST studies ranged from 20 to 50, whereas 103 participants completed the DST in this study.

The participants' indicated negative affect significantly increased in the DST and in all but one of the TSST studies (Table 2). The standardized effect sizes for the change in negative affect in the TSST studies ranged from 0.281 to 1.015, with a combined effect size of 0.667. The calculated effect for the increase in negative affect in the DST participants was 0.427.

The reported positive affect significantly increased in the DST participants in this study, whereas the results of the 4 TSST studies did not reveal significant changes in positive affect (Table 2). The standardized effect sizes for the change in positive affect in the TSST studies ranged from 0.022 to 0.363, with a combined effect size of 0.119. The calculated effect for the increase in positive affect in the DST participants was 0.382.

Table 2. Overview of studies used for meta-analytical comparison of the Digital Stress Test (DST) with the Trier Social Stress Test (TSST) effect including paired *t* test results for each study.

Study and PANAS[a] subscale	Baseline score, mean (SD)	Score after, mean (SD)	Change, mean (SD)	Paired *t* test		
				t test (*df*)	*P* value	Cohen d_z
DST (n=103)						
NA[b]	1.57 (0.56)	1.84 (0.70)	0.27 (0.61)	–4.51 (102)	*<.001* [c]	0.427
PA[d]	3.08 (0.65)	3.37 (0.85)	0.29 (0.61)	–4.84 (102)	*<.001*	0.382
TSST [37] (n=26)						
NA	1.28 (0.38)	1.82 (0.66)	0.55 (0.68)	–4.08 (25)	*<.001*	1.015
PA	2.95 (0.45)	2.94 (0.59)	–0.01 (0.51)	0.11 (25)	.91	0.022
TSST [34] (n=26)						
NA	1.33 (0.32)	1.66 (0.57)	0.32 (0.44)	–3.75 (25)	*<.001*	0.694
PA	2.95 (0.56)	2.97 (0.64)	0.02 (0.45)	–0.17 (25)	.86	0.026
TSST [36] (n=20)						
NA	1.36 (0.33)	1.51 (0.71)	0.16 (0.68)	–1.02 (19)	.32	0.281
PA	2.75 (0.42)	3.02 (0.96)	0.27 (0.82)	–1.47 (19)	.16	0.363
TSST [38] (n=50)						
NA	1.43 (0.56)	1.85 (0.72)	0.42 (0.53)	–5.64 (49)	*<.001*	0.655
PA	3.02 (0.57)	2.88 (0.68)	–0.14 (0.47)	2.10 (49)	.04	0.221

[a]PANAS: Positive and Negative Affect Schedule.
[b]NA: negative affect.
[c]Italics emphasize significance.
[d]PA: positive affect.

Discussion

Principal Findings

In this proof-of-concept study, we evaluated the feasibility of a fully digitalized acute stress paradigm for smartphones, the DST, to induce and record psychosocial stress responses in outside-the-laboratory settings. We compared it with a digital control condition (C-DST) in a large web-based study and set the effect size of the participants' indicated affect changes in the context of results previously achieved in the TSST. To our knowledge, this is the first study evaluating the stress reactivity of an experimenter-independent paradigm that does not include any human-human interaction.

We showed that the DST significantly induced higher levels of perceived stress and negative affect than the control condition. In addition to feeling more stressed, DST participants also reported similar increases in related affects such as frustration, shame, and overstrain. Notably, the reported increases in negative affect indicated by DST participants not only significantly exceeded those of participants performing the C-DST but were also comparable with those reported by TSST participants in previous studies regarding the calculated effect sizes.

These findings provide convincing evidence that an acute psychosocial stress response can be induced with a smartphone without any further equipment or experimenters taking part. In particular, the DST managed to induce subjective stress even if the social-evaluative threat and uncontrollability [25] of this study can be assumed to be weaker than in previous studies. TSST participants performed the paradigm in the laboratory and were administered physiological measurements and watched by several experimenters, whereas the DST and C-DST were mainly performed at home without any additional procedures or people present. Participants in the web-based study took <10 minutes for the whole paradigm and could cancel the study at any time by simply closing the browser. Nevertheless, the mere framing of social evaluation, a difficult mathematical task, and a free speech task in front of the smartphone camera were sufficient to elicit a psychological stress response.

These findings extend the results obtained in other studies analyzing the stress induction potential of less controlled and experimenter-dependent stress paradigms. Virtual reality versions of the TSST successfully elicit psychosocial stress responses using prerecorded [43], animated [44-47], or even nonhuman robot audiences [48]. However, these protocols still require experimenters to conduct the procedure.

Although previous studies have focused on the development of more immersive and convincing virtual realities to improve stress induction [49], our results indicate that the procedure might be simplified and spare human-human interaction. The recently investigated internet-delivered TSST has already shown that a significant stress response can be induced without direct person-to-person contact [16,17]. Our study supports these findings and further leads to the assumption that psychosocial stress can be induced without any live interaction.

Interestingly, in addition to evoking a significant level of perceived stress and negative affect, the DST also increased the participants' positive affect. Increases in positive affect have also been reported in other studies, including stress tests [36,37]. We assume that the increase in this study was caused by an end-of-study relief and self-selection bias. First, the participants in this web-based study knew that the performance test would end after the last questionnaire, whereas, in many other studies, experimental measures or interventions followed the stress paradigm [14,50,51]. Second, participants with a strong decrease in their positive affect might have cancelled the study because of the very low cancellation barrier.

For a more detailed investigation of the stress induction potential of our new paradigm, we also examined the elements implemented for stress induction in the DST regarding the participants' responses. Previous work has highlighted the impact of social evaluation and unpredictability on stress response. In particular, public speaking parts have been shown to induce stress in participants [25,52,53]. In our study, we found a strong increase in perceived stress throughout the Math-Task and a subsequent slight decrease over the Speech-Task. The results of the posttest questionnaires also indicate that the participants perceived the Math-Task as the more stressful task. In contrast to the TSST, the Speech-Task was the last part of the procedure, and the participants knew that the study would end afterward. Thus, the affect ratings might also have been influenced by the task order and end-of-study relief. Another reason for the lower stress induced by the Speech-Task might be that speaking to the front camera without any real social evaluation does not induce as much stress as that experienced in live experiments. Similarly, other paradigms that include a social-evaluative stressor without direct human interaction resulted in weaker stress responses [44,46,48]. In addition, despite receiving live feedback from the audio input, the participants might not talk or might skip the task as there is no real experimenter control. Furthermore, the participants in this study knew that their recordings would not be saved or watched.

For upcoming web-based studies, permanently saving the videos and the possibility that experimenters watch them might increase psychosocial stress. In addition, improving the credibility of the automated analysis through the implementation of more sophisticated adaptive feedback might lead to a stronger feeling of social evaluation. Another approach to strengthen the social-evaluative characteristics of the DST could be to implement a prerecorded or animated audience instead of displaying the participants' own video recordings. In addition, strengthening the social comparison characteristics of the paradigm through fabricated comparisons of the performance during the Speech-Task (similar to the Math-Task) might lead to a stronger psychosocial stress induction.

Web-Based Feasibility of the DST

The evaluation of the DST in a web-based study highlights the potential of this paradigm. Within 2 weeks, nearly 600 participants performed one of the versions, and almost 300 completed it. By contrast, a recent review evaluating 35 TSST studies showed that the average number of participants was 47, with only 1 study including >100 participants [54]. Campbell and Ehlert [55] evaluated 359 TSST and TSST-related articles and found only 6 studies that reported >100 participants, presuming many more laborious and time-consuming studies. Even in the recently proposed web-based TSST, experimenters and actors need to be present during the web-based videoconferencing session, and the still laborious procedure is stated as a limitation by the authors [16].

Another advantage of the DST procedure is its inclusiveness, allowing for participation from any location and in different conditions. However, the number and composition of the participants highly depend on the recruitment process. Many participants entered this study because of its announcement in a well-known German political podcast and university mailing lists, which might have led to age and educational background selection bias in our sample. The participants in this study were mainly younger and from higher educational backgrounds. Previous studies have shown differences in stress reactivity according to age and socioeconomic status, which need to be addressed when interpreting the findings of this study. In some studies, physiological stress responses to cognitive challenges were stronger in older and higher-educated individuals [56-58]. Nicolson et al [59] found stronger cortisol reactivity in younger individuals and no age-related differences in emotional responses to a speech task. According to Dickerson et al [25], cognitive testing may be more stressful for older adults with higher levels of education as they perceive a greater threat of negative social evaluation. Moreover, the average lower digital literacy of older adults [60] may even increase the stress response in older participants in a smartphone-based paradigm such as the DST. However, future studies should verify the stress induction potential for individuals of other ages and educational backgrounds.

In a web-based study without any direct supervision, it is crucial to ensure that the participants follow the correct procedure of the experiment. Therefore, the participants were automatically reminded to continue when they did not react during the tasks for a certain time. In addition, we logged the study progress and excluded participants who were extreme outliers with respect to the study duration. In the future, we plan to also analyze the video recordings regarding compliance and include more detailed live feedback.

The barrier for dropping out of this web-based study was much lower than that in laboratory or other live-contact settings. The participants could cancel the study at any time simply by closing the browser of their smartphones. Although it was ethically favorable that participants did not need to continue when they felt overwhelmed by the test situation, this also affected the outcome of the study. Many participants (324/547, 59.2%) dropped out even though the procedure took <10 minutes and no personal data were saved permanently. Most DST participants (89/300, 29.7%) had already cancelled during the introduction, which was not observed in the C-DST group. We assume that the higher cancellation rate in the beginning was caused by technological problems or privacy concerns related to the video recording in the DST.

For future versions, we plan to emphasize the high standard of data protection implemented in the DST and a cancellation procedure that allows for further decision-making regarding the submitted data and short feedback on the cancellation reasons.

Limitations

Previous studies have highlighted the long-term consequences of acute stress–induced physiological changes [61], which were not evaluated in this study. Although, in some experiments, correlations between psychological and physiological stress responses could be found [62,63], others could not verify this [25,55,64]. Hellhammer and Schubert [65] found that psychological measurements during, but not before or after, the TSST were related to physiological responses. The DST participants reported the highest level of perceived stress between the 2 tasks, indicating that physiological changes might also have taken place. Even if it is not yet clear whether the stress response elicited by the DST entails physiological changes, addressing psychological stress reactivity plays an important role in the individual quality of life [66] and mental well-being [67]. Previous studies have shown the effects of interventions on psychological well-being [68,69], which highlights the potential use of the DST for evaluating stress intervention strategies.

Nevertheless, the stress induction potential of the DST should be confirmed in a follow-up study including measurements of other stress-relevant systems such as the sympathetic nervous system and the hypothalamic-pituitary-adrenal axis [55,70].

Future Research

Several improvements to the stress induction procedure as well as the usability have been outlined. In particular, additional adaptive feedback algorithms that react to the participants' live-recorded behavior might improve the credibility of the social-evaluative framing and enhance compliance.

To further validate the DST, we plan to compare the psychological and physiological stress responses, including cortisol, heart rate, and blood pressure measurements, of participants undergoing the TSST and DST in a within-subject design. Next, to improve and validate the stress induction procedure, we aim to adjust and evaluate the video data collection in the DST and build a large data set of stress test videos.

The DST might then be easily applied to different (clinical) cohorts (eg, stress in patients with chronic pain [71], stress in patients with cancer [72], and stress in students [73]) and contexts (eg, job stress [74] and parental stress [75]) from any internet-connected location worldwide. In contrast to existing protocols, this would also allow for the conduction of stress studies in outside-the-laboratory scenarios and with individuals from diverse cultural, ethnic, and geographical backgrounds (eg, remote cultures) [76].

In contrast, the multimodal video data collected using the DST could serve as the basis for the development of video-based stress analysis algorithms using machine learning methods [77]. Baird et al [78] combined 3 data sets including videos and voice recordings of participants undergoing the TSST in separate studies for the prediction of acute stress responses. Consequently, the data obtained with the DST could enrich existing video data sets and be used in combination with them (eg, pretraining for personalized models [79] and cross-model transfer learning [78]) to improve the quality of the algorithms. From a more long-term perspective, these algorithms might be used within the DST to provide feedback on a participant's stress reactivity and evaluate personal prevention or intervention strategies (eg, resilience trainings [80]).

Conclusions

To the best of our knowledge, this is the first approach to a standardized digital stress paradigm that can be carried out using only a smartphone. Moreover, our results imply that psychosocial stress can be induced through cognitive-verbal performance tasks and additional framings in a fully automated web application.

The ability to conduct (stress) studies without any experimenter or additional equipment required can also be seen as a potential turning point for translating traditional (stress) research to the wild. Owing to the web application–based mobile architecture, future researchers can quickly prepare, conduct, adapt, and evaluate studies anywhere—including basic and clinical research. In accordance with the principles of open access, the source code of the DST and C-DST is publicly available, and both applications can be freely used for research purposes upon request.

Future studies will evaluate the potential of the implemented video recording capability to provide a high-quality stress data set for algorithm development. This study may serve as inspiration to bridge the gap between classic psychological research and interdisciplinary computer science.

Acknowledgments

This study was funded by the Federal Ministry of Education and Research (award 01IS20046). The Federal Ministry of Education and Research had no involvement in the study design; the collection, analysis, and interpretation of the data; the writing of the report; or the decision to submit the paper for publication.

Conflicts of Interest

FB reports grants from the German Federal Ministry of Education and Research, grants from the German Federal Ministry of Health, grants from the Berlin Institute of Health, personal fees from Elsevier Publishing and Medtronic, grants from the Hans

Böckler Foundation, other grants from the Robert Koch Institute, grants from the Einstein Foundation, and grants from Berlin University Alliance outside the submitted work.

Multimedia Appendix 1

Video illustrating the procedure for the Digital Stress Test web application used in this study. The video has been modified to 2x playback speed, and recorded sound has been removed for publication purposes. For proper inspection or reading of specific parts, the video can be paused. Screen recordings were done by a research assistant in May 2021. The mock-up design was provided by Vectorium - de.freepik.com. A presentation version of the most recent digital stress test version without permanent data saving can be found at www.digitalstresstest.org.
[MP4 File (MP4 Video), 60338 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Video illustrating the procedure for the Control - Digital Stress Test web application used in this study. The video has been modified to 2x playback speed, and recorded sound has been removed for publication purposes. For proper inspection or reading of specific parts, the video can be paused. Screen recordings were done by a research assistant in May 2021. The mock-up design was provided by Vectorium - de.freepik.com. A presentation version of the most recent control digital stress test version without permanent data saving can be found at www.digitalstresstest.org/control.
[MP4 File (MP4 Video), 44183 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Distribution of age (Figure S1) and educational background (Figure S2) across Digital Stress Test and Control - Digital Stress Test participants.
[PDF File (Adobe PDF File), 72 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Mixed-model ANOVA results for the comparison of participants' affective responses indicated in the different VAS and PANAS questionnaires at different time points. Second sheet displays post-hoc analyses for significant effects.
[XLS File (Microsoft Excel File), 20 KB-Multimedia Appendix 4]

References

1. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA 2007 Oct 10;298(14):1685-1687. [doi: 10.1001/jama.298.14.1685] [Medline: 17925521]
2. Dedovic K, D'Aguiar C, Pruessner JC. What stress does to your brain: a review of neuroimaging studies. Can J Psychiatry 2009 Jan;54(1):6-15. [doi: 10.1177/070674370905400104] [Medline: 19175975]
3. Chen X, Gianferante D, Hanlin L, Fiksdal A, Breines JG, Thoma MV, et al. HPA-axis and inflammatory reactivity to acute stress is related with basal HPA-axis activity. Psychoneuroendocrinology 2017 Apr;78:168-176 [FREE Full text] [doi: 10.1016/j.psyneuen.2017.01.035] [Medline: 28209543]
4. Wolf OT. Stress and memory retrieval: mechanisms and consequences. Curr Opin Behav Sci 2017 Apr;14:40-46. [doi: 10.1016/j.cobeha.2016.12.001]
5. von Dawans B, Strojny J, Domes G. The effects of acute stress and stress hormones on social cognition and behavior: current state of research and future directions. Neurosci Biobehav Rev 2021 Feb;121:75-88. [doi: 10.1016/j.neubiorev.2020.11.026] [Medline: 33301780]
6. Het S, Vocks S, Wolf JM, Herpertz S, Wolf OT. Treatment-resistant blunted HPA activity, but reversible cardiovascular stress reactivity in young women with eating disorders. Front Psychiatry 2020 Jul 22;11:726 [FREE Full text] [doi: 10.3389/fpsyt.2020.00726] [Medline: 32793011]
7. Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: a survey. J Biomed Inform 2019 Apr;92:103139 [FREE Full text] [doi: 10.1016/j.jbi.2019.103139] [Medline: 30825538]
8. Pisanski K, Kobylarek A, Jakubowska L, Nowak J, Walter A, Błaszczyński K, et al. Multimodal stress detection: testing for covariation in vocal, hormonal and physiological responses to Trier Social Stress Test. Horm Behav 2018 Nov;106:52-61. [doi: 10.1016/j.yhbeh.2018.08.014] [Medline: 30189213]
9. Mahesh B, Hassan T, Prassler E, Garbas JU. Requirements for a reference dataset for multimodal human stress detection. In: Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications Workshops. 2019 Presented at: PerComW '19; March 11-15, 2019; Kyoto, Japan p. 492-498. [doi: 10.1109/percomw.2019.8730884]
10. Picard RW. Automating the recognition of stress and emotion: from lab to real-world impact. IEEE MultiMedia 2016 Aug 5;23(3):3-7. [doi: 10.1109/mmul.2016.38]

11. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'—a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 1993;28(1-2):76-81. [doi: 10.1159/000119004] [Medline: 8255414]
12. Allen AP, Kennedy PJ, Dockray S, Cryan JF, Dinan TG, Clarke G. The Trier Social Stress Test: principles and practice. Neurobiol Stress 2017 Feb;6:113-126 [FREE Full text] [doi: 10.1016/j.ynstr.2016.11.001] [Medline: 28229114]
13. Birkett MA. The Trier Social Stress Test protocol for inducing psychological stress. J Vis Exp 2011 Oct 19(56):3238 [FREE Full text] [doi: 10.3791/3238] [Medline: 22042290]
14. Goodman WK, Janson J, Wolf JM. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. Psychoneuroendocrinology 2017 Jun;80:26-35. [doi: 10.1016/j.psyneuen.2017.02.030] [Medline: 28292684]
15. Kudielka BM, Hellhammer DH, Kirschbaum C. Ten years of research with the trier social stress test—revisited. In: Harmon-Jones E, Winkielman P, editors. Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior. New York, NY, USA: The Guilford Press; 2007:56-83.
16. Gunnar MR, Reid BM, Donzella B, Miller ZR, Gardow S, Tsakonas NC, et al. Validation of an online version of the Trier Social Stress Test in a study of adolescents. Psychoneuroendocrinology 2021 Mar;125:105111 [FREE Full text] [doi: 10.1016/j.psyneuen.2020.105111] [Medline: 33341502]
17. Eagle DE, Rash JA, Tice L, Proeschold-Bell RJ. Evaluation of a remote, Internet-delivered version of the Trier Social Stress Test. Int J Psychophysiol 2021 Jul;165:137-144. [doi: 10.1016/j.ijpsycho.2021.03.009] [Medline: 33865901]
18. Berretz G, Packheiser J, Kumsta R, Wolf OT, Ocklenburg S. The brain under stress-a systematic review and activation likelihood estimation meta-analysis of changes in BOLD signal associated with acute stress exposure. Neurosci Biobehav Rev 2021 May;124:89-99. [doi: 10.1016/j.neubiorev.2021.01.001] [Medline: 33497786]
19. Fehlner P, Bilek E, Harneit A, Böhringer A, Moessnang C, Meyer-Lindenberg A, et al. Neural responses to social evaluative threat in the absence of negative investigator feedback and provoked performance failures. Hum Brain Mapp 2020 Jun 01;41(8):2092-2103 [FREE Full text] [doi: 10.1002/hbm.24932] [Medline: 31958212]
20. Dedovic K, Renwick R, Mahani NK, Engert V, Lupien SJ, Pruessner JC. The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. J Psychiatry Neurosci 2005 Sep;30(5):319-325 [FREE Full text] [Medline: 16151536]
21. Schwabe L, Haddad L, Schachinger H. HPA axis activation by a socially evaluated cold-pressor test. Psychoneuroendocrinology 2008 Jul;33(6):890-895. [doi: 10.1016/j.psyneuen.2008.03.001] [Medline: 18403130]
22. Bali A, Jaggi AS. Clinical experimental stress studies: methods and assessment. Rev Neurosci 2015;26(5):555-579. [doi: 10.1515/revneuro-2015-0004] [Medline: 26020552]
23. Lejuez CW, Kahler CW, Brown RA. A modified computer version of the Paced Auditory Serial Addition Task (PASAT) as a laboratory-based stressor. Behav Ther 2003;26(4):290-293.
24. Renaud P, Blondin JP. The stress of Stroop performance: physiological and emotional responses to color–word interference, task pacing, and pacing speed. Int J Psychophysiol 1997 Sep;27(2):87-97. [doi: 10.1016/s0167-8760(97)00049-4]
25. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004 May;130(3):355-391. [doi: 10.1037/0033-2909.130.3.355] [Medline: 15122924]
26. Digital Stress Test. URL: https://www.digitalstresstest.org [accessed 2022-06-29]
27. Digital Stress Test Control. URL: https://www.digitalstresstest.org/control [accessed 2022-06-29]
28. Het S, Rohleder N, Schoofs D, Kirschbaum C, Wolf OT. Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. Psychoneuroendocrinology 2009 Aug;34(7):1075-1086. [doi: 10.1016/j.psyneuen.2009.02.008] [Medline: 19307062]
29. Wiemers US, Schoofs D, Wolf OT. A friendly version of the trier social stress test does not activate the HPA axis in healthy men and women. Stress 2013 Mar;16(2):254-260. [doi: 10.3109/10253890.2012.714427] [Medline: 22813431]
30. Digital Stress Test Code. URL: https://www.digitalstresstest.org/code [accessed 2022-06-29]
31. Lange K, Kühn S, Filevich E. "Just Another Tool for Online Studies" (JATOS): an easy solution for setup and management of Web servers supporting online studies. PLoS One 2015 Jun 26;10(6):e0130834 [FREE Full text] [doi: 10.1371/journal.pone.0130834] [Medline: 26114751]
32. Leiner DJ. SoSci Survey (Version 3.2.33). SoSci Survey. 2014. URL: https://www.soscisurvey.de [accessed 2021-02-01]
33. Thompson ER. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). J Cross Cult Psychol 2007 Mar 1;38(2):227-242. [doi: 10.1177/0022022106297301]
34. Herten N, Otto T, Wolf OT. The role of eye fixation in memory enhancement under stress - an eye tracking study. Neurobiol Learn Mem 2017 Apr;140:134-144. [doi: 10.1016/j.nlm.2017.02.016] [Medline: 28267591]
35. Lesage FX, Berjot S, Deschamps F. Clinical stress assessment using a visual analogue scale. Occup Med (Lond) 2012 Dec;62(8):600-605. [doi: 10.1093/occmed/kqs140] [Medline: 22965867]
36. Herten N, Otto T, Adolph D, Pause BM, Kumsta R, Wolf OT. Enhanced startle responsivity 24 hours after acute stress exposure. Behav Neurosci 2016 Oct;130(5):521-530. [doi: 10.1037/bne0000156] [Medline: 27362707]
37. Herten N, Pomrehn D, Wolf OT. Memory for objects and startle responsivity in the immediate aftermath of exposure to the Trier Social Stress Test. Behav Brain Res 2017 May 30;326:272-280. [doi: 10.1016/j.bbr.2017.03.002] [Medline: 28315373]

38. Wolf OT, Schulte JM, Drimalla H, Hamacher-Dang TC, Knoch D, Dziobek I. Enhanced emotional empathy after psychosocial stress in young healthy men. Stress 2015;18(6):631-637. [doi: 10.3109/10253890.2015.1078787] [Medline: 26365125]
39. Schmider E, Ziegler M, Danay E, Beyer L, Bühner M. Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. Methodology 2010 Jan;6(4):147-151. [doi: 10.1027/1614-2241/a000016]
40. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annu Rev Public Health 2002;23:151-169. [doi: 10.1146/annurev.publhealth.23.100901.140546] [Medline: 11910059]
41. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. Hoboken, NJ, USA: John Wiley & Sons; 2009.
42. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
43. Hawn SE, Paul L, Thomas S, Miller S, Amstadter AB. Stress reactivity to an electronic version of the Trier Social Stress Test: a pilot study. Front Psychol 2015 Mar 29;6:724 [FREE Full text] [doi: 10.3389/fpsyg.2015.00724] [Medline: 26074862]
44. Montero-López E, Santos-Ruiz A, García-Ríos MC, Rodríguez-Blázquez R, Pérez-García M, Peralta-Ramírez MI. A virtual reality approach to the Trier Social Stress Test: contrasting two distinct protocols. Behav Res Methods 2016 Mar;48(1):223-232. [doi: 10.3758/s13428-015-0565-4] [Medline: 25673321]
45. Fich LB, Jönsson P, Kirkegaard PH, Wallergård M, Garde AH, Hansen Å. Can architectural design alter the physiological reaction to psychosocial stress? A virtual TSST experiment. Physiol Behav 2014 Aug;135:91-97. [doi: 10.1016/j.physbeh.2014.05.034] [Medline: 24907691]
46. Shiban Y, Diemer J, Brandl S, Zack R, Mühlberger A, Wüst S. Trier Social Stress Test in vivo and in virtual reality: dissociation of response domains. Int J Psychophysiol 2016 Dec;110:47-55. [doi: 10.1016/j.ijpsycho.2016.10.008] [Medline: 27742258]
47. Jönsson P, Wallergård M, Osterberg K, Hansen AM, Johansson G, Karlson B. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: a pilot study. Psychoneuroendocrinology 2010 Oct;35(9):1397-1403. [doi: 10.1016/j.psyneuen.2010.04.003] [Medline: 20451329]
48. Turner-Cobb JM, Asif M, Turner JE, Bevan C, Fraser DS. Use of a non-human robot audience to induce stress reactivity in human participants. Comput Human Behav 2019 Oct;99:76-85. [doi: 10.1016/j.chb.2019.05.019]
49. Zimmer P, Buttlar B, Halbeisen G, Walther E, Domes G. Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. Psychoneuroendocrinology 2019 Mar;101:186-192. [doi: 10.1016/j.psyneuen.2018.11.010] [Medline: 30469086]
50. Dierolf AM, Schoofs D, Hessas EM, Falkenstein M, Otto T, Paul M, et al. Good to be stressed? Improved response inhibition and error processing after acute stress in young and older men. Neuropsychologia 2018 Oct;119:434-447. [doi: 10.1016/j.neuropsychologia.2018.08.020] [Medline: 30171874]
51. Merz CJ, Wolf OT, Schweckendiek J, Klucken T, Vaitl D, Stark R. Stress differentially affects fear conditioning in men and women. Psychoneuroendocrinology 2013 Nov;38(11):2529-2541. [doi: 10.1016/j.psyneuen.2013.05.015] [Medline: 23790683]
52. Sumter SR, Bokhorst CL, Miers AC, Van Pelt J, Westenberg PM. Age and puberty differences in stress responses during a public speaking task: do adolescents grow more sensitive to social evaluation? Psychoneuroendocrinology 2010 Nov;35(10):1510-1516. [doi: 10.1016/j.psyneuen.2010.05.004] [Medline: 20541871]
53. Westenberg PM, Bokhorst CL, Miers AC, Sumter SR, Kallen VL, van Pelt J, et al. A prepared speech in front of a pre-recorded audience: subjective, physiological, and neuroendocrine responses to the Leiden Public Speaking Task. Biol Psychol 2009 Oct;82(2):116-124. [doi: 10.1016/j.biopsycho.2009.06.005] [Medline: 19576261]
54. Narvaez Linares NF, Charron V, Ouimet AJ, Labelle PR, Plamondon H. A systematic review of the Trier Social Stress Test methodology: issues in promoting study comparison and replicable research. Neurobiol Stress 2020 Nov;13:100235 [FREE Full text] [doi: 10.1016/j.ynstr.2020.100235] [Medline: 33344691]
55. Campbell J, Ehlert U. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? Psychoneuroendocrinology 2012 Aug;37(8):1111-1134. [doi: 10.1016/j.psyneuen.2011.12.010] [Medline: 22260938]
56. Gotthardt U, Schweiger U, Fahrenberg J, Lauer CJ, Holsboer F, Heuser I. Cortisol, ACTH, and cardiovascular response to a cognitive challenge paradigm in aging and depression. Am J Physiol 1995 Apr;268(4 Pt 2):R865-R873. [doi: 10.1152/ajpregu.1995.268.4.R865] [Medline: 7733395]
57. Neupert SD, Miller LM, Lachman ME. Physiological reactivity to cognitive stressors: variations by age and socioeconomic status. Int J Aging Hum Dev 2006;62(3):221-235. [doi: 10.2190/17DU-21AA-5HUK-7UFG] [Medline: 16625938]
58. Seeman TE, Robbins RJ. Aging and hypothalamic-pituitary-adrenal response to challenge in humans. Endocr Rev 1994 Apr;15(2):233-260. [doi: 10.1210/edrv-15-2-233] [Medline: 8026389]
59. Nicolson N, Storms C, Ponds R, Sulon J. Salivary cortisol levels and stress reactivity in human aging. J Gerontol A Biol Sci Med Sci 1997 Mar;52(2):M68-M75. [doi: 10.1093/gerona/52a.2.m68] [Medline: 9060972]
60. Hargittai E, Piper AM, Morris MR. From Internet access to Internet skills: digital inequality among older adults. Univ Access Inf Soc 2018 May 3;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
61. Yaribeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: a review. EXCLI J 2017 Jul 21;16:1057-1072 [FREE Full text] [doi: 10.17179/excli2017-480] [Medline: 28900385]

62. Rimmele U, Zellweger BC, Marti B, Seiler R, Mohiyeddini C, Ehlert U, et al. Trained men show lower cortisol, heart rate and psychological responses to psychosocial stress compared with untrained men. Psychoneuroendocrinology 2007 Jul;32(6):627-635. [doi: 10.1016/j.psyneuen.2007.04.005] [Medline: 17560731]
63. Schlotz W, Kumsta R, Layes I, Entringer S, Jones A, Wüst S. Covariance between psychological and endocrine responses to pharmacological challenge and psychosocial stress: a question of timing. Psychosom Med 2008 Sep;70(7):787-796. [doi: 10.1097/PSY.0b013e3181810658] [Medline: 18725434]
64. Domes G, Heinrichs M, Rimmele U, Reichwald U, Hautzinger M. Acute stress impairs recognition for positive words--association with stress-induced cortisol secretion. Stress 2004 Sep;7(3):173-181. [doi: 10.1080/10253890412331273213] [Medline: 15764014]
65. Hellhammer J, Schubert M. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. Psychoneuroendocrinology 2012 Jan;37(1):119-124. [doi: 10.1016/j.psyneuen.2011.05.012] [Medline: 21689890]
66. Seo EJ, Ahn JA, Hayman LL, Kim CJ. The association between perceived stress and quality of life in university students: the parallel mediating role of depressive symptoms and health-promoting behaviors. Asian Nurs Res (Korean Soc Nurs Sci) 2018 Sep;12(3):190-196 [FREE Full text] [doi: 10.1016/j.anr.2018.08.001] [Medline: 30103040]
67. Strizhitskaya O, Petrash M, Savenysheva S, Murtazina I, Golovey L. Perceived stress and psychological well-being: the role of the emotional stability. In: Proceedings of the 7th Annual International Conference on Cognitive - Social, and Behavioural Sciences. 2018 Presented at: icCSBs '18; November 12-14, 2018; Moscow, Russia p. 162. [doi: 10.15405/epsbs.2019.02.02.18]
68. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013 Feb 08;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
69. Weiss LA, Westerhof GJ, Bohlmeijer ET. Can we increase psychological well-being? The effects of interventions on psychological well-being: a meta-analysis of randomized controlled trials. PLoS One 2016 Jun 21;11(6):e0158092 [FREE Full text] [doi: 10.1371/journal.pone.0158092] [Medline: 27328124]
70. Andrews J, Ali N, Pruessner JC. Reflections on the interaction of psychogenic stress systems in humans: the stress coherence/compensation model. Psychoneuroendocrinology 2013 Jul;38(7):947-961. [doi: 10.1016/j.psyneuen.2013.02.010] [Medline: 23522990]
71. McHugh RK, Kneeland ET, Edwards RR, Jamison R, Weiss RD. Pain catastrophizing and distress intolerance: prediction of pain and emotional stress reactivity. J Behav Med 2020 Aug;43(4):623-629 [FREE Full text] [doi: 10.1007/s10865-019-00086-5] [Medline: 31376099]
72. Powell N, Tarr A, Sheridan J. Psychosocial stress and inflammation in cancer. Brain Behav Immun 2013 Mar;30 Suppl:S41-S47. [doi: 10.1016/j.bbi.2012.06.015] [Medline: 22790082]
73. Travis F, Haaga DA, Hagelin J, Tanner M, Nidich S, Gaylord-King C, et al. Effects of Transcendental Meditation practice on brain functioning and stress reactivity in college students. Int J Psychophysiol 2009 Feb;71(2):170-176. [doi: 10.1016/j.ijpsycho.2008.09.007] [Medline: 18854202]
74. Limm H, Angerer P, Heinmueller M, Marten-Mittag B, Nater UM, Guendel H. Self-perceived stress reactivity is an indicator of psychosocial impairment at the workplace. BMC Public Health 2010 May 14;10:252 [FREE Full text] [doi: 10.1186/1471-2458-10-252] [Medline: 20470413]
75. Ostberg M. Parental stress, psychosocial problems and responsiveness in help-seeking parents with small (2-45 months old) children. Acta Paediatr 1998 Jan;87(1):69-76. [doi: 10.1080/08035259850157903] [Medline: 9510451]
76. Kirschbaum C. Why we need an online version of the Trier Social Stress Test. Psychoneuroendocrinology 2021 Mar;125:105129. [doi: 10.1016/j.psyneuen.2021.105129] [Medline: 33493741]
77. Goodday SM, Friend S. Unlocking stress and forecasting its consequences with digital technology. NPJ Digit Med 2019 Jul 31;2:75 [FREE Full text] [doi: 10.1038/s41746-019-0151-8] [Medline: 31372508]
78. Baird A, Triantafyllopoulos A, Zänkert S, Ottl S, Christ L, Stappen L, et al. An evaluation of speech-based recognition of emotional and physiological markers of stress. Front Comput Sci 2021 Dec 6;3:750284 [FREE Full text] [doi: 10.3389/fcomp.2021.750284]
79. Ghadiyaram D, Tran D, Mahajan D. Large-scale weakly-supervised pre-training for video action recognition. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019 Jun Presented at: CVPR '19; June 15-20, 2019; Long Beach, CA, USA p. 12038-12047 URL: https://ieeexplore.ieee.org/document/8953267/ [doi: 10.1109/cvpr.2019.01232]
80. Golden EA, Zweig M, Danieletto M, Landell K, Nadkarni G, Bottinger E, et al. A resilience-building app to support the mental health of health care workers in the COVID-19 era: design process, distribution, and evaluation. JMIR Form Res 2021 May 05;5(5):e26590 [FREE Full text] [doi: 10.2196/26590] [Medline: 33872189]

Abbreviations

API: application programming interface

C-DST: Control - Digital Stress Test
DST: Digital Stress Test
MIST: Montreal Imaging Stress Task
PANAS: Positive and Negative Affect Schedule
TSST: Trier Social Stress Test
VAS: Visual Analogue Scale

Edited by A Mavragani; submitted 28.07.21; peer-reviewed by R Ciorap, P Bhadra, H Mehdizadeh; comments to author 08.03.22; revised version received 24.05.22; accepted 26.05.22; published 15.07.22

Please cite as:
Norden M, Hofmann AG, Meier M, Balzer F, Wolf OT, Böttinger E, Drimalla H
Inducing and Recording Acute Stress Responses on a Large Scale With the Digital Stress Test (DST): Development and Evaluation Study
J Med Internet Res 2022;24(7):e32280
URL: https://www.jmir.org/2022/7/e32280
doi: 10.2196/32280
PMID:

Curriculum Vitae

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Komplette Publikationsliste

Norden, M., Hofmann, A. G., Meier, M., Balzer, F., Wolf, O. T., Böttinger, E., & Drimalla, H. (2022). Inducing and Recording Acute Stress Responses on a Large Scale With the Digital Stress Test (DST): Development and Evaluation Study. Journal of Medical Internet Research, 24(7), e32280. https://doi.org/10.2196/32280

Norden, M., Mayer, A., Molkenthin, D., Stubenvoll, C., & Drimalla, H. (2022). Validation of a Digital Stress Test (DST) for the induction of acute stress responses. https://doi.org/10.17605/OSF.IO/XWU9A

Norden, M., Wolf, O. T., Lehmann, L., Langer, K., Lippert, C., & Drimalla, H. (2022). Automatic Detection of Subjective, Annotated and Physiological Stress Responses from Video Data. 2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII), 1–8. https://doi.org/10.1109/ACII55700.2022.9953894

Kirchler, M., Konigorski, S., **Norden, M.**, Meltendorf, C., Kloft, M., Schurmann, C.& Lippert, C. (2022). transferGWAS: GWAS of Images Using Deep Transfer Learning”. Bioinformatics 38.14, pp. 3621–3628. doi: 10.1093/bioinformatics/btac369

Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich auf meinem Weg zur Fertigstellung dieser Dissertation unterstützt und begleitet haben. Zunächst möchte ich Herrn Prof. Dr. Dr. Balzer für die Möglichkeit, mich mit der Dissertation in dem interdisziplinären Feld zwischen Medizin, Psychologie und Informatik zu bewegen, danken. Seine Erfahrung und zügigen Rückmeldungen bei konkreten Fragen haben mich auf dem Weg zur Publikation und Dissertation sehr unterstützt.

Ich möchte auch allen Kolleg*innen und Ko-Autor*innen danken, die in dem Projekt an der praktischen Umsetzung der Studien, technologischen Implementierung und Auswertung beteiligt waren und durchgehend wertvolles Feedback gegeben haben. Meiner Familie danke ich dafür, dass sie auch auf dem etwas „unkonventionellen“ Weg der Forschung zu medizinischer Informatik (und nicht dem „klassischen“ Arzt-Weg) immer bedingungslos zu meiner Seite gestanden hat.

Insbesondere möchte ich meiner direkten Ansprechpartnerin und Betreuerin Prof. Dr. Hanna Drimalla Dank aussprechen. Über die gesamte Zeit, in allen Phasen der Dissertation, hatte sie ein offenes Ohr für mich und fand für meine (teils überlangen, überkritischen, teils über- oder demotivierten) Anregungen die richtigen Worte. Durch die Zusammenarbeit und das ständige „Ideen Ping-Pong“ mit ihr konnte ich unglaublich viel wissenschaftlich aber auch persönlich lernen. Die Arbeit in einer so lebhaften und herzlichen Atmosphäre hat letztlich dazu geführt, dass ich weiter mit viel Leidenschaft und Freude dem Weg der Forschung folge.

D a n k e