

Development and Validation of a Digital Ambulatory Stress Induction Toolbox: The Salzburg Mobile Stress Induction (SMSI)

Thomas Vikoler, Jasmina Dubravac, Heike Ganesch, Eva Traut-Mattausch

Submitted to: Journal of Medical Internet Research
on: April 10, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	26
Figures	27
Figure 1.....	28
Figure 2.....	29
Figure 3.....	30
Figure 4.....	31
Figure 5.....	32
Multimedia Appendixes	33
Multimedia Appendix 1.....	34
Multimedia Appendix 2.....	34
Multimedia Appendix 3.....	34
Multimedia Appendix 4.....	34
Multimedia Appendix 5.....	34
Multimedia Appendix 6.....	34
Multimedia Appendix 7.....	34
Multimedia Appendix 8.....	34
Multimedia Appendix 9.....	34
Multimedia Appendix 10.....	35

Development and Validation of a Digital Ambulatory Stress Induction Toolbox: The Salzburg Mobile Stress Induction (SMSI)

Thomas Vikoler¹ MSc; Jasmina Dubravac¹ MSc; Heike Ganesch¹ MSc; Eva Traut-Mattausch¹ Dr phil

¹Department of Psychology University of Salzburg Salzburg AT

Corresponding Author:

Eva Traut-Mattausch Dr phil
Department of Psychology
University of Salzburg
Hellbrunner Straße 34
Salzburg
AT

Abstract

Background: The use of digital technology enabled the investigation of stress in everyday life. However, ambulatory research depends on the natural occurrence of stressful situations while most standardized stress inductions rely on cost- and labor-expensive laboratory experiments, which are limited in their infrequent applicability.

Objective: We developed the Salzburg Mobile Stress Induction (SMSI), a newly conceptualized toolbox including six different stress-inducing paradigms (Matrices, Cube Net, Arithmetic, Number Series, Word Scramble, and Word Pair) and one control paradigm (Caesar Cipher) which are based on cognitive performance tests. These seven tests are intended to provide researchers with an open access, standardized method to repeatedly induce stress in an ambulatory setting.

Methods: We recruited university students from a local university and via a crowdsourcing platform for a preregistered ambulatory study. After an initial web-based survey, participants performed the seven SMSI tests in randomized order over 4 days on their smartphones via the m-Path app. By comparing the stress-inducing tests with the control test, we investigated changes in momentary negative and positive affect from baseline (t0) to between (t1) and after (t2) each test using the International Positive and Negative Affect Schedule Short Form.

Results: A total of 100 participants (60/100 women; mean age 24.43 years, SD=6.21; 69/100 local sample; 31/100 crowdsourcing sample) completed all seven tests of the SMSI. Participants' negative affect significantly increased during all six stress-inducing tests compared to the control test from t0 to t1 ($P < .001$) as well as to t2 ($P < .001$) with medium to large effect sizes ($\eta^2 = .10$ to $.30$). Post-hoc pairwise comparisons showed significant increases of negative affect during all stress-inducing tests from t0 to t1 ($P < .001$) and t2 ($P < .001$) and a slight increase in the control test from t0 to t2 ($P = .006$). Reported positive affect significantly differed between the stress-inducing tests and the control test from t0 to t1 ($P < .001$) and to t2 ($P < .001$) with medium to large effect sizes ($\eta^2 = .14$ to $.33$). Post-hoc pairwise comparisons revealed a significant increase in positive affect in the control test from t0 to t1 ($P < .001$) as well as to t2 ($P < .001$) and varying significant decreases to non-significant changes in the stress-inducing tests over time ($P > .99$ to $< .001$).

Conclusions: The SMSI presents novel and easy-to-implement standardized stress induction procedures to repeatedly evoke stress in ambulatory research. We discuss new opportunities for positive (eu-)stress inductions and outline subsequent validation studies combining physiological stress assessment and ambulatory methods. The development of additional language versions of the SMSI is illustrated.

(JMIR Preprints 10/04/2025:75785)

DOI: <https://doi.org/10.2196/preprints.75785>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org>, I will be able to access the full text of my article.

No. Please do not make my accepted manuscript PDF available to anyone.



Original Manuscript

Original Paper

Development and Validation of a Digital Ambulatory Stress Induction Toolbox: The Salzburg Mobile Stress Induction (SMSI)

Thomas Vikoler, University of Salzburg, Department of Psychology, Salzburg, Austria
Jasmina Dubravac, University of Salzburg, Department of Psychology, Salzburg, Austria
Heike Ganesch, University of Salzburg, Department of Psychology, Salzburg, Austria
Eva Traut-Mattausch, University of Salzburg, Department of Psychology, Salzburg, Austria

Corresponding Author:

Eva Traut-Mattausch

Department of Psychology

University of Salzburg,

Salzburg, 5020

Austria

Phone: +43 (0)662 / 8044 – 5149

Fax: +43 (0)662 / 8044 – 5126

Email: eva.traut-mattausch@plus.ac.at

Abstract

Background: The use of digital technology enabled the investigation of stress in everyday life. However, ambulatory research depends on the natural occurrence of stressful situations while most standardized stress inductions rely on cost- and labor-expensive laboratory experiments, which are limited in their infrequent applicability.

Objective: We developed the Salzburg Mobile Stress Induction (SMSI), a newly conceptualized toolbox including six different stress-inducing paradigms (Matrices, Cube Net, Arithmetic, Number Series, Word Scramble, and Word Pair) and one control paradigm (Caesar Cipher) which are based on cognitive performance tests. These seven tests are intended to provide researchers with an open access, standardized method to repeatedly induce stress in an ambulatory setting.

Methods: We recruited university students from a local university and via a crowdsourcing platform for a preregistered ambulatory study. After an initial web-based survey, participants performed the seven SMSI tests in randomized order over 4 days on their smartphones via the m-Path app. By comparing the stress-inducing tests with the control test, we investigated changes in momentary negative and positive affect from baseline (t0) to between (t1) and after (t2) each test using the International Positive and Negative Affect Schedule Short Form.

Results: A total of 100 participants (60/100 women; mean age 24.43 years, SD=6.21; 69/100 local sample; 31/100 crowdsourcing sample) completed all seven tests of the SMSI. Participants' negative affect significantly increased during all six stress-inducing tests compared to the control test from t0 to t1 ($P < .001$) as well as to t2 ($P < .001$) with medium to large effect sizes ($\eta^2 = .10$ to $.30$). Post-hoc pairwise comparisons showed significant increases of negative affect during all stress-inducing tests from t0 to t1 ($P < .001$) and t2 ($P < .001$) and a slight increase in the control test from t0 to t2 ($P = .006$). Reported positive affect significantly differed between the stress-inducing tests and the control test from t0 to t1 ($P < .001$) and to t2 ($P < .001$) with medium to large effect sizes ($\eta^2 = .14$ to $.33$). Post-hoc pairwise comparisons revealed a significant increase in positive affect in the control test from t0 to t1 ($P < .001$) as well as to t2 ($P < .001$) and varying significant decreases to non-significant changes in the stress-inducing tests over time ($P > .99$ to $< .001$).

Conclusions: The SMSI presents novel and easy-to-implement standardized stress induction procedures to repeatedly evoke stress in ambulatory research. We discuss new opportunities for positive (eu-)stress inductions and outline subsequent validation studies combining physiological stress assessment and ambulatory methods. The development of additional language versions of the SMSI is illustrated.

Keywords: stress induction; ambulatory stress induction; ambulatory assessment; acute stress; negative affect; positive affect; cognitive stress test; within-person; smartphone; digital health; mobile health; open access material.

Introduction

Relevance and Rationale

Stress is an everyday experience, which can fluctuate throughout the day and across days, weeks, months, years, across the lifetime. The effects of stress on health have been thoroughly researched: Stress has been shown to be a major cause for cardiovascular diseases, diabetes, dysregulation of the immune system, cognitive impairments, and mood disorders [1–3]. To assess and provide measures to cope with stress in everyday life, research has increasingly moved away from laboratory, cross-sectional or continual panel studies and more towards an ambulatory methodological approach to capture the momentary experience or to deliver interventions in everyday life, for example with ecological momentary assessments [4] or ecological momentary interventions [5], respectively. However, ambulatory methods are limited in their ability to optimally process stress as they need to

operate at the right time when stress occurs or is likely to occur. To circumvent this difficulty, researchers conventionally apply stress induction paradigms in an experimental setting that provides a controlled basis to examine acute stress responses or to evaluate the effectiveness of stress management interventions.

Current (Digital) Stress Paradigms and Their Limitations

There are many different stress induction methods, which can be organized into three categories: Physical, physiological, and psychological (for an overview see [6]). Most of the current stress paradigms have in common that they are performed in the laboratory to facilitate standardization at the expense of transferability to everyday life. Most of them are time and cost intensive, require direct participant-investigator interactions within artificial laboratory settings, can only be repeated at long intervals and are limited in their application to small sample sizes. To investigate acute stress responses in everyday life in a standardized way, validated stress paradigms are needed which can be integrated into a digital environment to overcome these limitations.

The most popular stress paradigm and considered as the gold standard in experimental research is the Trier Social Stress Test (TSST) [7]. The TSST procedure includes three sections: A preparation and delivery section of a mock job interview as well as an arithmetic calculation section, which are carried out in front of research assistants disguised as evaluators to induce psychosocial stress. The TSST has been shown multiple times to effectively induce stress (see meta-analysis [8]). To reduce the resource intensiveness and variability of the procedure, the TSST protocol was adapted into the virtual environment by utilizing avatars or pre-recorded videos of real confederates (see meta-analysis [9]). This virtual version of the TSST showed similar yet less intense stress responses than the traditional TSST for which the level of immersivity played a role (ie, low immersive screen-based environments vs. high immersive 3D projections or head mounted devices that replace all visual and audio cues). However, the virtual TSST has mostly been tested in laboratories, and when applied remotely, they still required live participant-investigator interactions, while at the same time significantly taking participants out of their everyday lives. For example, a recent study [10] applied the virtual TSST in everyday life via Zoom. Participants had to be at home in an undisturbed room, while simultaneously a researcher was still required to instruct and monitor the participants during the entire procedure. Furthermore, the effectiveness of both the traditional as well as the virtual TSSTs decreased with multiple exposures at short time intervals indicating habituation effects [11], which limits repeated examinations of stress responses in everyday life. Although, virtual stress paradigms underline the relevance of standardized examinations of acute stress responses in everyday life, they mirror laboratory experiments and do not represent ambulatory stress inductions in everyday life.

The Digital Stress Test (DST) [12] recently introduced a first prototype that allows for remotely inducing stress without the need for communication with researchers in a standardized way. The DST included tasks similar to the TSST and provided a port to the ambulatory setting using smartphone-based procedures of an arithmetic calculation task and a free speech task. They included social evaluative threat by directly displaying the participant's performance compared to other participants (calculation task) as well as by video recording and displaying of the participant's face during the participation (both tasks). The DST demonstrated the potential of ambulatory stress paradigms outside of an artificial laboratory setting. The authors found a significant increase in negative affect after completing the entire DST compared to their neutral control test (a procedure resembling the stress test with less stressful framings and tasks). When examining the participants' experience of stress throughout the DST, its performance test characteristics and the arithmetic calculation task were identified as most stressful elements while the speech task and (audio-)visual features of both DST subtests were reported to not elevate stress that much. Since the DST is a newly developed tool, insights into effects of repeated applications are currently not available. When comparing similar

tests like the Montreal Imaging Stress Test (which the arithmetic calculation task of the DST has partly adapted) [13], one study using the version for repeated application found indications of an anticipation and consequently a habituation effect within the same test sessions [14]. Nonetheless, the DST showed that cognitive performance tests can be a feasible way to induce stress in everyday life. Ambulatory research in particular can benefit greatly from the possibilities of repeated stress inductions in everyday life, for which we present a toolbox of multiple newly developed cognitive performance tests that operate on a standardized digital procedure.

The Salzburg Mobile Stress Induction (SMSI)

With the Salzburg Mobile Stress Induction (SMSI) we introduce our newly conceptualized and developed toolbox for researchers to remotely induce stress in everyday life. The SMSI comprises a variety of distinct stress paradigms to repeatedly evoke acute stress responses in the participants' familiar environment and simultaneously circumvent habituation effects [11]. Additionally, the SMSI can operate solely digitally without the need of any direct contact between the participant and the study coordinator—from installation (eg, web browser-based) to execution (eg, app-based via m-Path [15]).

Concept of the SMSI Tests

The SMSI contains a total of seven different cognitive tests—six stress-inducing tests and one neutral control test. All seven tests are based on and thus framed as cognitive intelligence tests for the application in everyday life. Alongside a variety of tests, we wanted to ensure that these also differed in terms of content. Two of the six stress-inducing tests each comprise figural (Matrices, Cube Net), numeric (Arithmetic Task; Number Series), and word-based tasks (Word Scramble, Word Pairs) while the seventh neutral control test was conceptualized as a word-based test (Caesar Cipher). The three word-based tests we present in this study are in the German language. Test material for the English language version has already been developed and is currently in its validation process (T Vikoler, G, Biczkowski, L Brandt, M Brunner, E Traut-Mattausch, unpublished data, 2025). Example screenshots of SMSI test tasks as well as applied feedback are shown in Figure 1 and in Figures S1 to S4 in Multimedia Appendix 1. Following, we describe the shared layout of all seven tests. Varying characteristics of the individual stress induction tests as well as the neutral control test are detailed in the respective sections below. Each test consists of 2-3 blocks with 12 single-choice tasks each. To answer the single-choice questions four options were presented with the fourth stating that none of the three answers are correct. Participants had a constant time limit of 12 seconds per questions, except for the Word Pair test, which included a one-minute memory phase followed by a five second response time per question. After each individual question immediate feedback is given: If the question was answered incorrectly or too late, prominent negative feedback will be displayed in form of a bright red screen with a big white cross in the middle (Figure 1e). In contrast, if answered correctly, only a small green tick is shown on a white background (Figure 1f).

Matrices (MT)

The Matrices paradigm (MT) is introduced as a test of perceptual and logical thinking skills and is based on commonly used intelligence matrix tests (eg, Raven Progressive Matrices [16]). Each task comprises of a 3x3 grids filled with figures and shapes with one field left blank (Figure 1a). The objective is to deduce which figure would logically complete the pattern, based on structural rules of the matrix. These rules may apply either horizontally, vertically, and/or diagonally within the matrix. With increasing difficulty, the designs get more and more elaborate and intricate as well as the logical rules increase in complexity (eg, rotation rules with different rotation degrees). Further, conditional rules were introduced and combined with other rules (eg, presence or absence of patterns in one field determine the shading or position within the field of a preceding element) to enhance the difficulty level. In addition to the increasingly difficult rules and patterns in the depicted matrix and

in the answer options, the position of empty fields within the matrix changes over the course of the test to increase unpredictability.

Cube Net (CN)

The Cube Net paradigm (CN) is introduced as a test of spatial imagination and logical thinking skills and is based on spatial ability tests (eg, [17]). In this test a two-dimensional cube net of a three-dimensional cube is presented, and the objective is to determine which of the presented three-dimensional cubes could be constructed by folding the given net (Figure 1b). The cube nets include varying patterns and forms (eg, geometrical forms or shadings) in varying positions. With increasing difficulty, the designs on the cubes' sides get more intricate and include more complex asymmetrical patterns or even identical sides. These configurations demand higher cognitive processing and challenge the spatial visualization ability by requiring more mental rotations of the cubes.

Arithmetic (AR)

The Arithmetic paradigm (AR) is introduced as a test of mathematical skills. Each question displays two equations involving addition and subtraction of numbers between 1 and 9, separated by a horizontal line (Figure 1c). The objective is to solve each equation separately and then subtract the numerically smaller result from the larger one. Which of the two equations result in the larger score varies throughout the test. The displayed answer options range from single- to double-digits. With the increase in difficulty, the number of summands and subtrahends in the arithmetic equations increase progressively from two to six operators. The simultaneous calculation of up to 14 numbers in two equations and the memorization of the intermediate results contribute to the overall cognitive load.

Number Series (NS)

The Number Series paradigm (NS) is introduced as a test of the ability to deduce the rules of and continue numerical sequences and is based on common intelligence tests (eg, part of the intelligence structure test 2000 R [18]). In this test a series of numbers is presented, and the objective is to identify the rule of the sequence and determine the next number (Figure 1d). The array of numbers ranges from four to seven figures and includes single-, double- and triple-digit numbers. The tasks apply different principles: Additions, subtractions, multiplications, prime numbers, cross sums, exponentiates, and Fibonacci sequences. As the difficulty progresses, the complexity of the applied principles not only increases, but the principles also become interwoven (eg, subtract prime numbers) to challenge the cognitive processing and pattern recognition ability.

Word Scramble (WS)

The Word Scramble paradigm (WS) is introduced as a test of the ability to solve word puzzles and of word recognition skills. The WS test is the first of two tests of the SMSI, which is based on a pseudo intelligence test. Google searches for the imperative keywords "word scramble" in combination with "intelligence test" or "IQ" revealed 2.600 and 102.000 results respectively, illustrating the fundamental perception of word scramble tasks as equivalent to intelligence tests in the broad population. The object of this test is to unravel the displayed words with intermingled letters and identify the correct first letter of the sought word (Figure S1 in Multimedia Appendix 1). With increasing difficulty, the length of the words increased from 4-letter words up to 13-letter words. Additionally, difficulty increases by incorporating more complex words including repetition and/or double occurrences of vowels or consonants. Moreover, a more specialized and technical vocabulary is introduced to further enhance the challenge of word recognition.

Word Pair (WP)

The Word Pair paradigm (WP) is introduced as a test of the memory capacity and is based on widely used instruments for assessing explicit episodic memory like the Verbal Paired Associates test (subtest of the Wechsler Memory Scale [19]). The object is to memorize within 1 minute the presented list of word pairs (Figure S2 in Multimedia Appendix 1). Afterwards any first word from one of the previously presented word pairs is displayed and the task is to select the corresponding

matching word (Figure S3 in Multimedia Appendix 1). After six questions, a new word pair list is introduced, with a total of four lists. To bring the overall completion time of the WP in line with the other SMSI tests due to the memorization phases and to increase the level of difficulty, the time to answer each question was shortened (5 seconds). To increase the difficulty over time several adjustments were scheduled: First, the length of the word pair lists increases while the memorization time remains the same. Second, the utilized words get more complex, more technical and less familiar terms are chosen (eg, concrete vs. abstract words) which require more intricate and elaborate associations between the paired words. Third, the difficulty level increased with the structural level of the memorized and questioned word pairs as word pairs with or without similar prefixes or alliterations as well as word pairs with and without semantic relationships are used. Fourth, words from both previous word pair lists and questioned words are incorporated in subsequent word pair lists and questions to increase interference effects and to decrease the ability to distinguish and recall the correct second words.

Neutral Control Test – Caesar Cipher (CC)

The Caesar Cipher paradigm (CC) is introduced as a test of the ability to encrypt texts and was designed to serve as a neutral non-stress inducing control tests for the comparison against the stress-inducing tests of the SMSI. The CC is the second test of the SMSI, which is not based on scientific-based intelligence tests and instead depicts a commonly known encryption technique, namely the Caesar Cipher technique. Similarly to word scrambles, encryption tests are commonly perceived as an intelligence test equivalent, resulting in 34.900 prompts in a Google search for the mandatory keywords “encrypt” and “intelligence test” or 331.000 prompts for a pairing with the keyword “IQ”. The general design of the CC mirrored that of the experimental tests, including the procedure, design, number of questions and time limit. However, the level of difficulty does not increase beyond a moderate solvable level by employing a more straightforward rule-based approach. Contrary to mostly known cipher tasks found on the internet in our test the objective is not to decipher words with interchanged letters and find the correct sought word. The task is to encrypt a given word according to the Caesar cipher encryption code, by moving each letter down the alphabet by a set number of positions (eg, with an encryption code of 2 the letter “A” changes into a “C” or the letter “B” changes into “D”; Figure S4 in Multimedia Appendix 1). The low difficulty compared to the stress-inducing paradigms is achieved by reducing the mental steps to arrive at the correct response to a minimum. This was done by adapting the response options. The displayed possible decrypted words resembled closely, with only a maximum of two letters being differed. This intentional design enables easy-to-apply strategies of focusing on the only varying parts of the response options and the corresponding letters of the original word instead of moving every letter of the word down the alphabet. With easier tasks the time limit is consequently less demanding, resulting in better results with less arousing negative feedback. However, to maintain the cover of a performance intelligence test an increase in difficulty was mandatory. For this purpose, the words to be encoded become gradually longer from three letters up to six letters.

SMSI Test Material Availability

The SMSI presents a new tool for researchers to implement stress induction paradigms in their research, be it in laboratories or in ambulatory studies. To enable easy utilization of the SMSI and in accordance with the principles of open access, the entire test materials of the SMSI are publicly available at the Open Science Framework [20]. In addition, we publicly shared preset structures for the immediate import of the tests in m-Path [15]. Test materials of the English version are available on request.

Development and Procedure of the SMSI

Building on results of the DST [12], which identified participation in performance tests, time

restriction, uncontrollability, immediate feedback, and task difficulty as one of the most stress-inducing elements, we developed seven independent tests designed as quasi-intelligence test tasks. During the development of the SMSI, we drew on various sources to conceptualize varying stress-inducing types of performance tests. An arithmetic calculation type task was directly inspired by the DST and adapted to match the other tests outline. Commonly used intelligence test types (eg, matrices) were adduced in the development process. Other task types were derived from internet research on cognitive tasks (eg, word scramble). In addition to investigating and reviewing of possible stress-inducing tasks, we sought for a cognitive performance task which could provide low to mediocre difficulty to serve as a neutral control test by also not being easily identified as such.

Each test starts off with a short introduction and one example item to ensure comprehension of the objective of the respective task. Difficulty in the stress-inducing tests increased rapidly during the first block: The first four questions were designed to be easier to solve, followed by four moderate to difficult questions and four highly difficult questions at the end of the first block. In the second block of the stress-inducing tests, only the first four questions were designed as medium to difficult questions, whereas the remaining questions remained highly difficult. SMSI tests with three blocks continued the difficulty level of the end of the second block. Contrary, the control test only increases the difficulty of its questions to a mediocre level across all blocks. The difficulty was complemented with a time restriction for each question. Three of the seven SMSI tests (AR, WS, and CC) included three instead of two blocks of tasks. When administering multiple or all SMSI tests repeatedly within a short period of time, the different number of task blocks create a certain feeling of unpredictability in terms of the end of the tasks. Figure 2 gives an overview of the structure of the SMSI tests. After each separate task, we built in enhanced negative feedback on incorrect or overdue answers and simpler and more negligible positive feedback on correct answers, which should amplify the feeling of failure and consequently of stress. The respective feedback is only displayed for 1 second, after which the next question starts forthwith. An example of a complete smartphone procedure of the Arithmetic test of the SMSI can be seen in Multimedia Appendix 2.

Objectives and Hypotheses

The aim of this study was to validate the SMSI as a toolbox for repeatedly inducing stress digitally in everyday life. Further, we wanted to evaluate the SMSI's feasibility and stress-inducing potential against a neutral control test based on the same procedure principals of the other cognitive performance tests. We preregistered this study at the Open Science Framework [22] where we hypothesized that, negative affect scores increase significantly more in the experimental tests (MT, CN, NS, AR, WS, and WP) compared to the control test (CC) from baseline (before the first task block t0) to after the first task block (t1) and to after the second task block (t2) (Figure 2). In addition to assessing the changes in negative affect, we also investigated the effect of the seven SMSI tests on the participants' positive affect responses.

Methods

Pilot Study

To test the fundamental operability of the SMSI's smartphone procedure in conjunction with the used m-Path app [15], we conducted a pilot study with 3 participants who performed all seven tests over four days. We adjusted usability issues and revised for comprehensive m-path installation as well as test introductions.

Evaluation of the SMSI

To examine the feasibility and stress-inducing potential of the SMSI, we first conducted a longitudinal web- and smartphone-based study. In this study, participants performed all seven SMSI tests while surveying their affective states.

Participants and Recruitment

Overview

Participants consisted of only university students and were recruited in two ways: First, by distributing the link of the web-based initial survey within the local university network via social networks (eg, Facebook and WhatsApp), university e-mail lists and in university courses (via QR code and mail lists). Second, by distributing the link of another web-based initial survey (see *Framing* section) via crowdsourcing. A total of 229 participants from the local university subsample (164/229, 71.6%) and the crowdsourcing subsample (65/229, 28.4%) completed the initial survey, registered in m-Path [15] and were eligible for the smartphone study procedure. The seven SMSI tests were randomized within each participant and completed over four consecutive days (two tests per day; Figure 3). Eventually, of the 229 participants, 137 (57.8%) did not complete all SMSI tests and were therefore excluded from the analyses.

Framing and Incentive

The study was framed as “a study about the validation of a new smartphone-based tool for testing cognitive intelligence in everyday life via numerous cognitive tests”. For their participation in this study bachelor psychology students at the University of Salzburg were incentivized with university participation credits whereas crowdsourcing participants were remunerated for participating in the initial survey via the crowdsourcing platform as well as for participating each test. In the crowdsourcing subsample, the initial survey was framed differently—as a “study about personality, self-perception, and smartphone usage during their studies”—due to payment procedure incompatibility of both the initial survey part and the smartphone study part. At the end of this initial survey, the crowdsourcing participants were then invited to take part in the subsequent study, which continued with the same intelligence test cover story. During the smartphone procedure each SMSI test was introduced separately at the beginning to test an explicit skill (eg, mathematical, spatial imagination or logical thinking skills).

Study Procedure

The entire study procedure spanned six to seven days. Participants completed the initial survey in a web browser which included the cover story (see *Framing* section) and information about the study procedure, informed consent, demographic questions and trait questionnaires. At the end of the initial survey, participants received a step-by-step image-based instruction (Multimedia Appendix 3) on how to download the m-Path app on their own smartphone device, register and connect to the research account. One to two days later the smartphone study procedure started within the m-Path app (depending on the time of registering by the participant and implementation of the study by the investigators within the m-Path app).

The app-based study procedure lasted for five days. On the first day, participants received two notifications on their smartphone as a reminder that the study is about to start on the following morning (between 9 AM and 6 PM). On the following four days, participants received two notifications per day (one in the morning at 9 AM and in the afternoon at 3 PM) to participate in an intelligence test. Each test could only be started once and was open for three hours. The tests were randomized within the participants using block randomization. On the fifth day in the afternoon prompt, instead of a test participants were asked to answer feedback questions regarding the completed tests which concluded the study. Figure 3 shows an exemplary schedule of the smartphone study procedure.

Participants from the local university subsample were debriefed after the completion of the data collection. Participants from the crowdsourcing subsample were linked at the end of the final survey to an external survey regarding their payment information and afterwards debriefed. All participants could quit the study any time by closing (and deinstalling) the app.

m-Path App-Based Study Data Collection

The app-based data collection started with Day 2 of the smartphone study procedure (Figure 3). Each test procedure started with a short introduction, followed by instructions for the respective test and a simple example task, supplemented by a detailed description of how to solve the example task. We then assessed the baseline (t0) of the participants' momentary negative (5 items; Cronbach α =0.77 to 0.88) and positive affect (5 items; Cronbach α =0.87 to 0.93) with the International Positive and Negative Affect Schedule Short Form (PANAS [21]). Detailed Cronbach α of both subscales for each test and measurement time are shown in Table S1 in Multimedia Appendix 4. Additionally, two more items were presented at the same time regarding the participants' feelings of stress and frustration. These questions were rated on a visual analogue scale ranging from 0 (*not at all*) to 100 (*extremely*). Afterwards the first block of the respective SMSI test started. After completion of the first block, momentary affect was measured again (t1) in the same way as at baseline. Next came the second block of tasks, followed by a third assessment of momentary affect (t2). In the CC, AR, and WS tests, a third block of tasks and a fourth assessment of momentary affect were carried out. Each SMSI test concluded with feedback questions regarding the performed test (eg, comprehensible instructions, difficulty of the tasks, distraction during the tasks, and more or less time required). A detailed design of the app-based study procedure and data collection can be found in Figure 2.

Statistical Analysis

Power

In their study, Norden et al. [12] showed that the math test of the DST had a group $\times$ time interaction effect of $\eta_p^2=0.05$. Using this score within G*Power (Version 3.1.9.7) in an a priori sample size calculation resulted in an effect size $f=0.23$. With this effect size score, seven groups and three repeated measurements, the sample size power analysis revealed a minimum required sample size of 91. In the end, we gathered data from 100 participants who performed all seven tests of the SMSI.

Statistical Analysis

Two two-way repeated measures ANOVAs using SPSS (Version 29.0.0.0) were conducted to analyze the participants' change in negative and positive affect mean scores of the PANAS [21] with the within-factors *test* (SMSI stress-inducing tests MT, CN, NS, AR, WS, and WP against the neutral control test CC) and *time* (t0 against t1 and t2). Contrasts were calculated for test (stress-inducing tests MT, CN, NS, AR, WS, and WP against the neutral control test CC) and for time (t0 vs. t1 and t0 vs. t2).

Due to their robustness [23] against non-normal distributed data and violations of sphericity in conjunction with a large sample size given in our study, we conducted the analyses as scheduled using the recommended Greenhouse-Geisser corrected degrees of freedom for all models. To test the change in affect scores of each SMSI test separately, Bonferroni adjusted post-hoc pairwise comparisons were conducted. We calculate separate models to control for the between-subject factors sample and gender differences in changes of affective responses.

Ethics Approval

The study procedures, informed consent, data processing, debriefing and withdrawal from the study were conducted in accordance with the General Data Protection Regulation and ethics approval for the study was granted by the ethics committee of the University of Salzburg (EK-GZ: 34/2014). The data processing by m-Path used in this study is in accordance with the General Data Protection Regulation. Participants had the option of manually deleting their online data and blocking the research account in the app at any time.

Results

Participants and Completion Rate

In total, 229 university students registered for the study in m-Path with their smartphones of which 100 (43.7%) participants completed all seven SMSI tests (69/100, local university subsample; 31/100, crowdsourcing subsample). A detailed description of the compliance frequencies and ratio from study registration to completion from the total and both subsamples can be found in Table 1. For all SMSI tests, there was a 0% (0/229) dropout rate during the test procedures, that is, every participant who started a test and took part in a baseline assessment also completed the respective test as well as each subsequent assessment. Overall, the sample comprised 60 women, 34 men, and 1 person who identified as non-binary. Demographic data from 5 participants could not be assigned to smartphone data due to mismatching between the individually generated participant codes in the initial survey and in the m-Path registration. The participants were on average 24.43 years old ($SD=6.21$). A detailed description of the socio-demographic information of the total sample and split for both subsamples can be found in Table S2 in Multimedia Appendix 5. On average, the participants completed each test procedure in 11.34 minutes ($SD=7.54$). A detailed description of the required times for each test can be found in the Table S3 in Multimedia Appendix 6).

Table 1. Compliance frequencies and rate in the total sample and both subsamples. Percentages refer to the number of participants registered for the study in the m-Path app in each (sub-)sample.

Amount of Completed Tests	Total sample, n(%)	Local university subsample, n(%)	Crowdsourcing subsample, n(%)
0 ^a	229 (100%)	164 (100%)	65 (100%)
1	206/229 (90.0%)	153/164 (93.3%)	53/65 (81.5%)
2	200/229 (87.3%)	149/164 (90.9%)	51/65 (78.5%)
3	192/229 (83.8%)	144/164 (87.8%)	48/65 (73.8%)
4	182/229 (79.5%)	135/164 (82.3%)	47/65 (72.3%)
5	169/229 (73.8%)	125/164 (76.2%)	44/65 (67.7%)
6	147/229 (64.2%)	105/164 (64.0%)	42/65 (64.6%)
7	100/229 (43.7%)	69/164 (42.1%)	31/65 (47.7%)

^aParticipants who registered for the study in the m-Path app.

Negative Affect

The two-way repeated measures ANOVA regarding participants' change in negative affect showed significant main effects for the factors *test* ($F_{5.51,545.48}=5.13$, $P<.001$, $\eta_p^2=.05$) and *time* ($F_{1.18,116.62}=107.50$, $P<.001$, $\eta_p^2=.52$) as well as a significant *test*×*time* interaction effect ($F_{9.04,895.14}=7.96$, $P<.001$, $\eta_p^2=.07$). Tests of within-subject contrasts showed that negative affect scores significantly changed from t0 to t1 as well as to t2 in each stress-inducing test (MT, CN, AR, NS, WS, and WP) compared to the CC control test with medium to large effect sizes (Table 2). Post-hoc pairwise comparisons for each test separately (Table 3) showed that negative affect scores significantly increased from t0 to t1 as well as from t0 to t2 in all tests except in the CC control test in which there was only a significant increase from t0 to t2. There were significant increases in negative affect from t1 to t2 in the MT, NS, WS, WP and CC tests but not in the CN and AR tests. Changes in negative affect are shown in Figure 4. When controlling for *sample* (local university and crowdsourcing subsample; Table S4 in Multimedia Appendix 7) and *gender* (female and male; Table S5 in Multimedia Appendix 8), both show no significant main or interaction effects on changes in

negative affect scores by the SMSI tests. Thus, even controlling for *sample* or *gender*, our results showed that all stress-inducing tests of the SMSI robustly induced stress by eliciting negative affect, confirming our preregistered hypothesis.

Table 2. Contrast statistics of the two-way repeated measures ANOVA for negative affect measured by of the Positive and Negative Affect Schedule (PANAS) of the stress-inducing tests against the control test of the Salzburg Mobile Stress Induction (SMSI) from baseline (t0) to after the first task block (t1) and after the second task block assessments (t2; N=100). MT=Matrices; CN=Cube Net; AR=Arithmetic; NS=Number Series; WS=Word Scramble; WP=Word Pair; CC (CT)=Caesar Cipher (control test); η_p^2 =partial eta squared.

CC vs. (CT)	t0 vs. t1			t0 vs. t2		
	<i>F</i> test (1, 99)	<i>P</i> value	η_p^2	<i>F</i> test (1, 99)	<i>P</i> value	η_p^2
MT	29.02	<.001 ^a	.23	42.06	<.001	.30
CN	15.36	<.001	.13	10.63	.002	.10
AR	29.07	<.001	.23	26.23	<.001	.21
NS	20.98	<.001	.18	34.83	<.001	.26
WS	32.25	<.001	.25	33.67	<.001	.25
WP	15.31	<.001	.13	16.94	<.001	.15

^aItalics emphasize significance.

Table 3. Mean and SD of negative affect measured by of the Positive and Negative Affect Schedule (PANAS) as well as Bonferroni adjusted *P* values of the post-hoc pairwise comparisons between baseline (t0), after the first task block (t1) and after the second task block assessments (t2) for each test of the Salzburg Mobile Stress Induction separately and multivariate *F* statistics of time (t0, t1, and t2) within each test. MT=Matrices; CN=Cube Net; AR=Arithmetic; NS=Number Series; WS=Word Scramble; WP=Word Pair; CC (CT)=Caesar Cipher (control test); η_p^2 =partial eta squared.

Test	Negative affect								
	t0, mean(SD)	t1, mean(SD)	t2, mean(SD)	t0 vs. t1, <i>P</i> value	t1 vs. t2, <i>P</i> value	t0 vs. t2, <i>P</i> value	<i>F</i> test (2, 98)	<i>P</i> value	η_p^2
MT	11.12 (12.72)	21.01 (18.35)	25.02 (19.80)	<.001 ^a	<.001	<.001	45.94	<.001	.48
CN	16.38 (18.14)	22.76 (20.75)	24.08 (21.04)	<.001	.296	<.001	17.19	<.001	.26
AR	15.20 (15.62)	25.49 (20.39)	27.28 (21.27)	<.001	.170	<.001	33.07	<.001	.40
NS	11.04 (12.94)	19.93 (18.21)	25.59 (19.23)	<.001	<.001	<.001	43.54	<.001	.47
WS	10.66 (12.27)	19.91 (17.48)	22.58 (20.41)	<.001	.006	<.001	33.69	<.001	.41
WP	12.35 (13.64)	19.59 (17.10)	22.50 (18.56)	<.001	.004	<.001	20.34	<.001	.29
CC (CT)	15.30 (15.07)	16.54 (15.67)	18.52 (16.74)	.631	.044	.006	5.82	.004	.11

^aItalics emphasize significance.

Positive Affect

When examining the effects of the SMSI tests on positive affect, we found significant main effects of *test* ($F_{5,20,515.04}=5.76$, $P<.001$, $\eta_p^2=.06$) and *time* ($F_{1,26,124.74}=4.82$, $P=.022$, $\eta_p^2=.05$) as well as a significant *test*×*time* interaction effect ($F_{8,40,831.45}=8.08$, $P<.001$, $\eta_p^2=.08$). Tests of within-subject contrasts showed significant differences in the change of positive affect over time in all stress-inducing tests (MT, CN, AR, NS, WS, and WP) when compared to the CC control test with medium to large effect sizes (Table 4). Post-hoc pairwise comparisons for each test separately (Table 5) showed that positive affect scores significantly increased in the CC control test and significantly decreased in the WS test from t0 to t1. Positive affect scores significantly increased in the CC control test and significantly decreased in the MT and WS tests (trend in the AR test) from t0 to t2 as well as significantly decreased in the MT, AR and the WP tests from t1 to t2. Changes in positive affect are illustrated in Figure 5. When controlling for *sample* (local university and crowdsourcing subsample; Table S6 in Multimedia Appendix 9), we found a significant main effect of *sample* but no significant interactions with *test* or *time* concluding in no deviation of the underlying effects. When controlling for *gender* (female and male; Table S7 in Multimedia Appendix 10), we found no significant main or interaction effect of *gender* on changes in positive affect scores by the SMSI tests. While the SMSI stress-inducing tests partially decreased participants' positive affect, participants reported increasing positive affect when taking the CC control test, when controlled for both *sample* or *gender*.

Table 4. Contrast statistics of the two-way repeated measures ANOVA for positive affect measured by the Positive and Negative Affect Schedule of the stress-inducing tests against the control test of the Salzburg Mobile Stress Induction from baseline (t0) to after the first task block (t1) and after the second task block assessments (t2; N=100). MT=Matrices; CN=Cube Net; AR=Arithmetic; NS=Number Series; WS=Word Scramble; WP=Word Pair; CC (CT)=Caesar Cipher (control test); η_p^2 =partial eta squared.

CC (CT) vs.	t0 vs. t1			t0 vs. t2		
	<i>F</i> test (1, 99)	<i>P</i> value	η_p^2	<i>F</i> test (1, 99)	<i>P</i> value	η_p^2
MT	20.22	<.001 ^a	.17	41.57	<.001	.30
CN	19.37	<.001	.16	23.05	<.001	.19
AR	16.05	<.001	.14	35.74	<.001	.27
NS	15.50	<.001	.14	19.06	<.001	.16
WS	38.08	<.001	.23	48.12	<.001	.33
WP	18.09	<.001	.15	25.67	<.001	.21

^aItalics emphasize significance.

Table 5. Mean and SD of negative affect subscale measured by the Positive and Negative Affect Schedule (PANAS) and Bonferroni adjusted *P* values and effect sizes (η_p^2) of the post-hoc pairwise comparisons between baseline (t0), after the first task block (t1) and after the second task block assessments (t2) for each test of the Salzburg Mobile Stress Induction separately and multivariate *F* statistics of time (t0, t1, and t2) within each test. MT=Matrices; CN=Cube Net; AR=Arithmetic; NS=Number Series; WS=Word Scramble; WP=Word Pair; CC (CT)=Caesar Cipher (control test); η_p^2 =partial eta squared.

Test	Positive affect								
	t0, mean(S D)	t1, mean(S D)	t2, mean(S D)	t0 vs. t1, <i>P</i> value	t1 vs. t2, <i>P</i> value	t0 vs. t2, <i>P</i> value	<i>F</i> test (2, 98)	<i>P</i> value	η_p^2
MT	53.60	52.83	48.40	>.99	<.001	.002	8.29	<.001	.15

	(23.97)	(23.82)	(23.40)		^a				
CN	48.76 (22.10)	47.79 (22.08)	46.91 (22.66)	>.99	.873	.700	0.77	.466	.02
AR	49.79 (23.25)	49.59 (23.72)	46.59 (23.26)	>.99	.003	.099	5.90	.004	.11
NS	52.43 (22.79)	51.78 (23.55)	50.78 (24.32)	>.99	.751	.710	0.97	.384	.02
WS	54.32 (24.82)	50.88 (53.58)	49.60 (24.23)	.004	.409	.001	6.88	.002	.12
WP	56.29 (21.61)	56.11 (20.10)	53.38 (22.01)	>.99	.032	.236	3.67	.039	.06
CC (CT)	51.61 (20.48)	58.11 (20.37)	58.81 (20.86)	<.001	>.99	<.001	14.11	<.001	.22

^aItalics emphasize significance.

Discussion

The purpose of this research was to present the development and validation of new stress induction paradigms. With the introduction of the SMSI, we provide a new toolbox to induce acute stress repeatedly in everyday life. In this preregistered study, we evaluated the potential effectivity and feasibility of the SMSI to remotely induce stress in an ambulatory, digital framework via smartphone application. Starting from the initial survey and installation, a large sample of participants performed the seven SMSI tests over four days in their everyday lives, without the need for a laboratory setting or participant-investigator interactions. We compared the effectiveness of the six stress-inducing tests (MT, CN, AR, NS, WS, and WP) with the similar designed control test (CC).

Principal Results

We found support for our hypothesis that participants perceived significant more negative affect after completing the stress-inducing tests compared to the control test with medium (CN) to large (MT, AR, NS, WS, and WP) effect sizes. Participants reported the strongest increase of negative affect after the first task block of the stress-inducing tests and it continued to increase significantly in the MT, NS, WS, and WP tests. In the CN and AR tests, the increase in negative affect leveled off after the first task block, yet participants' negative affect was still significantly higher after the second task block compared to the baseline. Although, the CC test slightly increased negative affect over time, this increase occurred later in the procedure which can be attributed to exposure to an ongoing participation in a performance test. Overall, the effects of the CC test on negative affect are minimal compared to the stress-inducing tests. The MT test showed the strongest stress inducing potential, followed by the NS, WS, AR, WP and CN tests.

All in all, the SMSI provides a feasible way to induce acute stress responses outside the laboratory which is simple to implement in various smartphone applications (eg, m-Path[15]). These findings outline the potential of cognitive tasks as stress induction paradigms without components of social-evaluative threat, such as the TSST [7] or the DST [12] applied. This corresponds with the findings of the latter that in the study on the DST social comparison characteristics played a lesser role in the stress-inducing process. The incremental value of the SMSI lies in its potential for repetitive usage, as it offers six different stress paradigms with comparable effectiveness to its mobile stress-inducing predecessor, the DST. The SMSI reduces risks of habituation effects [11] and enables researchers to tailor various stress inductions to their specific needs in their ambulatory studies (eg, ecological momentary assessments [4] or ecological momentary interventions [5]) by freely choosing from the

provided tests of the toolbox.

Difficult demanding tasks—be it demands in everyday life or in our case by taking the SMSI tests—do not only lead to an increased negative stress experience it may affect positive experiences as well [24]. Our results showed that participants' positive affect changed significantly different when performing the stress-inducing tests compared to when they performed the neutral control test. In four of the six stress-inducing tests (MT, AR, WS, and WP) positive affect significantly decreased over the course of the procedure. Although it has been theorized that demanding tasks can enhance positive outcomes (eg, positive emotions) [24], in one momentary assessment study, the presence of stressors (eg, heightened job, private life or study-related demands) was related to lower positive stress responses [25]. Contrary to the stress-inducing tests of the SMSI, participants reported an increase of positive affect after performing the CC control test similar to the findings of the DST [12]. Although the CC test was established as a control test for negative affect and negative stress these results might indicate that the CC test could represent a new type of a stress paradigm: A positive, eustress-inducing paradigm. Eustress (and positive affect as its indicator) arises when a stressor is appraised less as an hindrance and more as a challenge, which leads to a feeling of accomplishment (for an overview of the concept of eustress see [26]). This experience of eustress might be the product of completing and solving the CC test, as it was designed to be manageable. These results open new possibilities for positive psychology research. There are far fewer positive than there are negative mood induction procedures in the literature [27] and to our knowledge, this is the first method using cognitive tasks with the potential to elevate not only positive affect but also eustress by applying challenging yet solvable tasks. However, the CC test (and potential adapted tests) still requires further experiments with dedicated eustress markers (eg, eustress specific questionnaires [25]) to provide convincing evidence of a eustress induction.

Feasibility of the SMSI in Ambulatory Research

The main objective of the SMSI is to provide feasible stress induction paradigm to repeatedly implement in ambulatory studies. Therefore, its feasibility in such research is of crucial importance. This study evaluated the SMSI's applicability within its intended ambulatory framework.

Similar to many other ambulatory research projects, compliance to study prompts posed certain challenges in this study. Many participants who initially registered for this study (129/229, 56.3%) missed at least one mobile prompt to participate in a scheduled test. Repeated measurements and study interactions are prone to increase participant burden due to their time-consuming nature and their inherent interference with participants' daily lives [28]. As a result, several meta-analyses [29–33] investigated characteristics of compliance in ambulatory assessment. In three meta-analyses [30,32,33] participants' compliance was positively related to age, which could explain some of the compliance rate in this study as the sample comprised of mainly younger adults, although other findings reported no such association [29,31]. Stress inductions are inherently intensive, something researchers should consider when applying these in ambulatory studies. According to Ottenstein and Werner [29], compliance rate is lower in intervention studies than in studies without interventions. In addition, a recent study evaluating the effects of intensity in ambulatory assessment studies showed that the length of the questionnaire (30 vs. 60 item questionnaires) was positively related to participant burden and negatively to compliance [34]. Although compliance presents a considerable challenge when it comes to ambulatory research, there are methods to increase consecutive participation. For example, pronounced reminders (eg, alarms or multiple prompts) can effectively increase compliance [33] as the main reason participants give for not responding to prompts was rarely a lack of motivation, but mainly that they were unable to respond at the appointed time [34]. In this study, we employed only one reminder 15 minutes after the prompt. In future studies, we plan to increase the number of reminders to further increase the participation likelihood.

Despite participants having the opportunity to cancel a test procedure at any time by closing the app,

all participants who initiated a test completed the respective tests to the end. In contrast, the DST [12] took less time than the SMSI tests, though over half of the initial participants of the study on the DST dropped out during the test procedure. Even a recent meta-analysis [29] showed that the percentage of dropouts were higher in ambulatory studies with interventions than in studies without. Although the compliance rate during this study's procedure presented similar challenges as in traditional ambulatory assessment studies, the 0% dropout rate during the test administration emphasizes the SMSI's practicality in ensuring psychological stress responses in ambulatory study designs.

Limitations and Future Research

Acute stress affects the entire central nervous system directly or indirectly but specific areas of the brain are critical keystones of the stress system [35]. As such, the hypothalamic-pituitary-adrenal axis functions as the major neuroendocrine stress axis of the human body, leading to an increased output of cortisol in stressful situations [36]. Two meta-analyses have already shown that the TSST increased the activity of the hypothalamic-pituitary-adrenal axis and consequently cortisol levels in traditional [8] and virtual environments [9], respectively. To provide convincing evidence of the SMSI potential to induce an acute stress response not only at the behavioral but also at the endocrinological level, we plan a follow-up study to investigate the SMSI's potential effects on physiological stress markers (eg, cortisol and alpha-amylase [37]). The collection of saliva samples in the participants' natural environment provides a non-invasive and relatively simple procedure with the possibility of measuring salivary cortisol repeatedly in short time intervals [38]. To combine the SMSI procedure with ambulatory cortisol monitoring, the number of tests would have to be changed to one per day (instead of two as in this study) to circumvent diurnal variations in cortisol levels [38]. To reduce the burden on participants, a scaled-down laboratory experiment beforehand could provide insight into the offset and peak of endocrinologic markers over time, allowing us to minimize the amount of saliva samples needed.

The SMSI allows to conduct stress inductions in diverse populations due to 4 tests (MT, CN, AR, and NS) are inherently nonverbal. However, the three word-based tests (WS, WP, and the control test CC) depend on language comprehension skills and are currently only validated in the German language. As a first step, we have developed an English version of these tests in order to be able to offer the SMSI more globally and are currently validating the entire SMSI in a population of native English speakers. The overall procedure of the tests mirrored the German study. In order to ensure comparable effects, the tasks of the word-based tests were reconstructed according to exactly the same patterns as the original tasks, taking into account the respective population and language comprehension. Further adaptations in additional languages would also enable standardized investigations into cross-cultural differences of acute stress responses [27] outside the laboratory.

Previous stress induction paradigms that use cognitive stimuli, such as the Montreal Imaging Stress Task [13] in laboratory settings or the arithmetic calculation task of the DST [12] in ambulatory applications, present the participants with their performance rate combined with a display of the average (mostly artificially preset) performance rate of potential peers to increase the effectiveness of their induction through social evaluative threat. The SMSI operates solely on cognitive stressors which has the advantage of less complex programming and simpler implementation in various software. Both approaches, cognitive stressors with or without social components, can allow for different areas of applications in research on everyday life. Cognitive demands in particular have increased in the course of digitalization in the world of work and in everyday life [39]. The boundary between work and private life became blurred, the demands on digital work evolved and the need for corresponding digital skills rose to meet these demands [40]. These new cognitive demands pose new challenges in the daily lives and depending on the specific cognitive demand, they can have negative or even positive effects [39]. Due to the cognitive nature of the tasks of SMSI, the toolbox

would therefore have a high practical relevance for investigating acute negative (and potential positive) stress responses to cognitive stressors in the natural environment of everyday life. Nevertheless, the potential of an extended version of SMSI that includes social comparisons should be investigated in future studies to explore the respective effects and interplay between the two mechanisms of cognitive stress stimuli and social evaluative threat.

Conclusions

The purpose of this research was to develop and validate a toolbox with newly conceptualized stress paradigms to ambulatorily evoke acute stress responses in study participants in their everyday lives without participant-investigator interactions. The advantage of the SMSI lies in its ability to repeatedly apply stress inductions to enable a standardized investigation of consecutive stress reactions outside the laboratory. This study provided first promising results on the potential of cognitive tasks to induce psychological stress in participants' natural environment using smartphone applications.

In addition to the effectiveness of the six stress-inducing tests and the feasibility of the SMSI, the potential of the CC test to elicit positive emotions offers new possibilities for stress research. By making the SMSI and its test materials open access (see SMSI Test Material Availability section), researchers can easily integrate and adapt mobile stress induction in their (ambulatory) studies as required. Future studies will further investigate the potential of both negative distress and positive eustress inductions in ambulatory research as well as the effect of the SMSI's on the physiological stress system.

Acknowledgements

Open access publication supported by the University of Salzburg Publication Fund.

Data Availability Statement

The SMSI test materials, datasets, and data processing scripts are openly available in the Open Science Framework depository [20].

Conflicts of Interest

None declared.

Abbreviations

AR: Arithmetic test
CC: Caesar Cipher test
CN: Cube Net test
DST: Digital Stress Test
MT: Matrices test
NS: Number Series test
PANAS: Positive and Negative Affect Schedule
WP: Word Pair test
WS: Word Scramble test
SMSI: Salzburg Mobile Stress Induction
TSST: Trier Social Stress Test

References

1. Fleshner M, Frank M, Maier SF. Danger Signals and Inflammasomes: Stress-Evoked Sterile Inflammation in Mood Disorders. *Neuropsychopharmacol* 2017 Jan;42(1):36–45. doi: 10.1038/npp.2016.125

2. McEwen BS. Physiology and Neurobiology of Stress and Adaptation: Central Role of the Brain. *Physiological Reviews* 2007 Jul;87(3):873–904. doi: 10.1152/physrev.00041.2006
3. Zafar MS, Nauman M, Nauman H, Nauman S, Kabir A, Shahid Z, Fatima A, Batool M. Impact of Stress on Human Body: A Review. *EJMED* 2021 May;3(3):1–7. doi: 10.24018/ejmed.2021.3.3.821
4. Shiffman S, Stone AA, Hufford MR. Ecological Momentary Assessment. *Annu Rev Clin Psychol* 2008 Apr;4(1):1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415
5. Heron KE, Smyth JM. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British J Health Psychol* 2010 Feb;15(1):1–39. doi: 10.1348/135910709X466063
6. Brunyé TT, Goring SA, Navarro E, Hart-Pomerantz H, Grekin S, McKinlay AM, Plessow F. Identifying the most effective acute stress induction methods for producing SAM- and HPA-related physiological responses: a meta-analysis. *Anxiety, Stress, & Coping* 2024 Jul;0(0):1–23. doi: 10.1080/10615806.2025.2450620
7. Kirschbaum C, Pirke K-M, Hellhammer D. The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology* 1993 Feb 1;28:76–81. doi: 10.1159/000119004
8. Gu H, Ma X, Zhao J, Liu C. A meta-analysis of salivary cortisol responses in the Trier Social Stress Test to evaluate the effects of speech topics, sex, and sample size. *Comprehensive Psychoneuroendocrinology* 2022 May;10:100125. doi: 10.1016/j.cpnec.2022.100125
9. Helminen EC, Morton ML, Wang Q, Felver JC. A meta-analysis of cortisol reactivity to the Trier Social Stress Test in virtual environments. *Psychoneuroendocrinology* 2019 Dec;110:104437. doi: 10.1016/j.psyneuen.2019.104437
10. Harvie HMK, Jain B, Nelson BW, Knight EL, Roos LE, Giuliano RJ. Induction of acute stress through an internet-delivered Trier Social Stress Test as assessed by photoplethysmography on a smartphone. *Stress* 2021 Nov;24(6):1023–1032. doi: 10.1080/10253890.2021.1995714
11. Kothgassner OD, Goreis A, Glenk LM, Kafka JX, Pfeffer B, Beutl L, Kryspin-Exner I, Hlavacs H, Palme R, Felnhofer A. Habituation of salivary cortisol and cardiovascular reactivity to a repeated real-life and virtual reality Trier Social Stress Test. *Physiology & Behavior* 2021 Dec;242:113618. doi: 10.1016/j.physbeh.2021.113618
12. 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 Jul;24(7):e32280. doi: 10.2196/32280
13. 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. *Journal of Psychiatry and Neuroscience* 2005 Sep;30(5):319–325. PMID:16151536

14. De Calheiros Velozo J, Vaessen T, Pruessner J, Van Diest I, Claes S, Myin-Germeys I. The repeated Montreal Imaging Stress Test (rMIST): Testing habituation, sensitization, and anticipation effects to repeated stress induction. *Psychoneuroendocrinology* 2021 Jun;128:105217. doi: 10.1016/j.psyneuen.2021.105217
15. Mestdagh M, Verdonck S, Piot M, Niemeijer K, Tuerlinckx F, Kuppens P, Dejonckheere E. m-Path: An easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. OSF. Preprint posted online January 2022. doi: 10.31234/osf.io/uqdfs
16. Raven J, Raven J. Raven Progressive Matrices. In: McCallum RS, editor. *Handbook of Nonverbal Assessment* Boston, MA: Springer; 2003. p. 223–237. doi: 10.1007/978-1-4615-0153-4_11
17. Pawlak-Jakubowska A, Terczyńska E. Evaluation of STEM students' spatial abilities based on a novel net cube imagination test. *Sci Rep Nature Publishing Group*; 2023 Oct;13(1):17296. doi: 10.1038/s41598-023-44371-5
18. Liepmann D, Beauducel A, Brocke B, Amthauer R. *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R) [intelligence structure test 2000 R]*. 2nd, expanded and revised ed. Hogrefe; 2007. Available from: <https://www.hogrefe.com/at/shop/intelligenz-struktur-test-2000-r.html> [accessed Sep 25, 2024]
19. Uttl B, Graf P, Richter LK. Verbal Paired Associates tests limits on validity and reliability. *Archives of Clinical Neuropsychology* 2002 Aug;17(6):567–581. doi: 10.1016/S0887-6177(01)00135-4
20. Vikoler T, Ganesch H, Dubravac J, Biczkowski G, Brandt L, Brunner M, Pletzer B, Traut-Mattausch E. The Salzburg Mobile Stress Induction (SMSI): Online Depository. Open Science Framework OSF; 2025 Apr 10; Available from: <https://osf.io/7hkju/>
21. Thompson ER. Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology* SAGE Publications Inc; 2007 Mar;38(2):227–242. doi: 10.1177/0022022106297301
22. Vikoler T, Ganesch H, Dubravac J, Traut-Mattausch E. Preregistration of a Smartphone-based Digital Toolbox to Induce Stress in Everyday Life – An Ecological Momentary Stress Induction [Preregistration]. OSF; 2023. Available from: <https://osf.io/bt7z5>
23. Blanca MJ, Alarcón R, Arnau J, García-Castro J, Bono R. How to proceed when normality and sphericity are violated in the repeated measures ANOVA. *Anales de Psicología / Annals of Psychology* Universidad de Murcia; 2024 Jul;40(3):466–480. doi: 10.6018/analesps.594291
24. Bakker AB, Demerouti E, Sanz-Vergel A. Job Demands–Resources Theory: Ten Years Later. *Annu Rev Organ Psychol Organ Behav* 2023 Jan 23;10(1):25–53. doi: 10.1146/annurev-orgpsych-120920-053933
25. Vikoler T, Kovács D, Traut-Mattausch E. The Di-Eu-Stress State Scale (DESS Scale) - Development and Validation of a Scale measuring State Distress and Eustress. *European Journal of Psychological Assessment* 2024 May; doi:

<https://doi.org/10.1027/1015-5759/a000829>

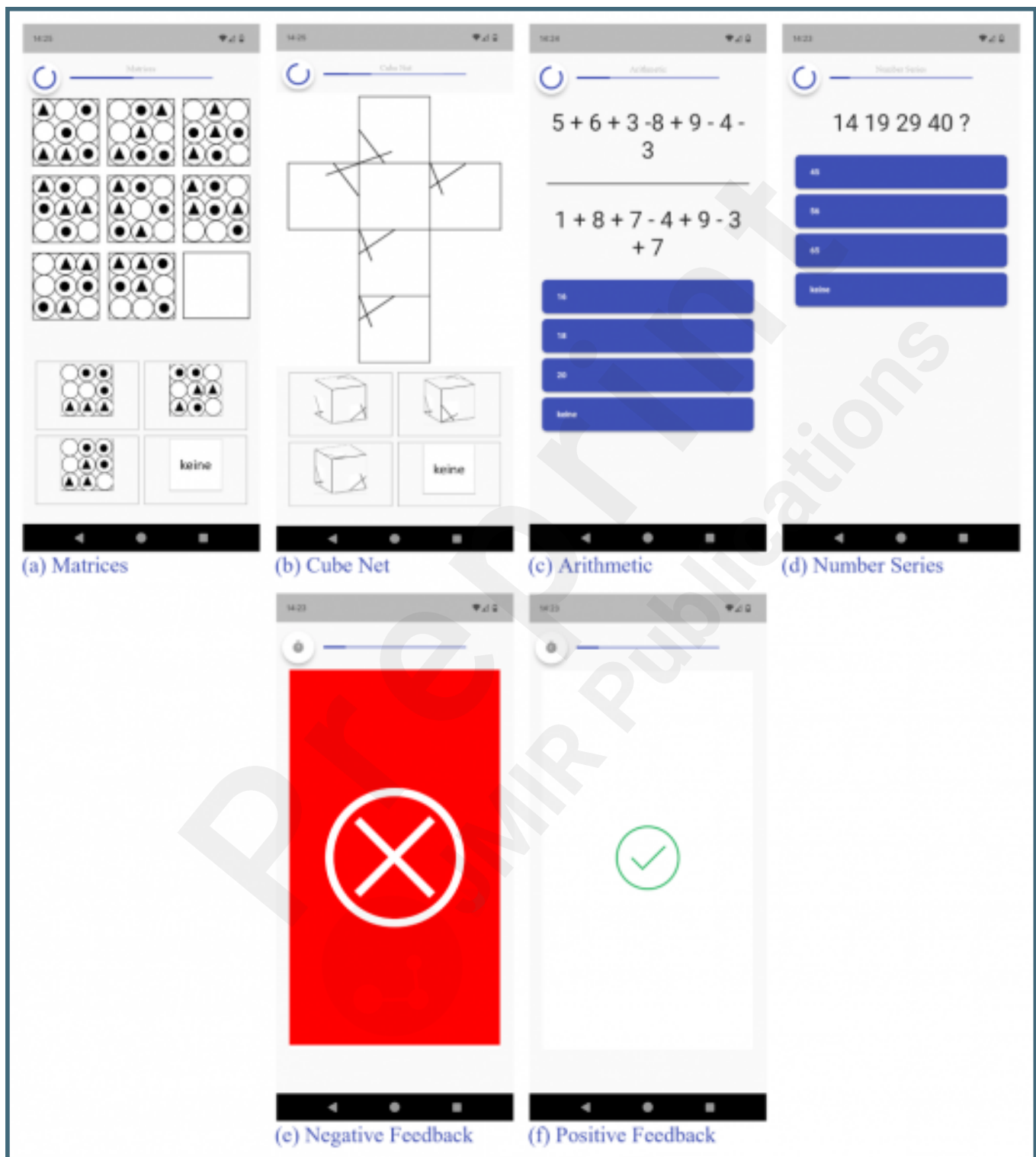
26. Simmons BL, Hanna A, Hu J. The holistic model of stress: Savoring eustress while coping with distress. *Handbook of occupational health psychology*, 3rd ed Washington, DC, US: American Psychological Association; 2024. p. 65–84. doi: 10.1037/0000331-004
27. Joseph DL, Chan MY, Heintzelman SJ, Tay L, Diener E, Scotney VS. The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin* 2020 Apr;146(4):355–375. doi: 10.1037/bul0000224
28. Carpenter RW, Wycoff AM, Trull TJ. *Ambulatory Assessment: New Adventures in Characterizing Dynamic Processes*. Assessment SAGE Publications Inc; 2016 Aug 1;23(4):414–424. doi: 10.1177/1073191116632341
29. Ottenstein C, Werner L. *Compliance in Ambulatory Assessment Studies: Investigating Study and Sample Characteristics as Predictors*. Assessment SAGE Publications Inc; 2022 Dec;29(8):1765–1776. doi: 10.1177/10731911211032718
30. Reiter T, Schoedel R. Never miss a beep: Using mobile sensing to investigate (non-)compliance in experience sampling studies. *Behav Res* 2024 Jun;56(4):4038–4060. doi: 10.3758/s13428-023-02252-9
31. Wrzus C, Neubauer AB. *Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields*. Assessment SAGE Publications Inc; 2023 Apr;30(3):825–846. doi: 10.1177/10731911211067538
32. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment US: American Psychological Association*; 2019;31(2):226–235. doi: 10.1037/pas0000662
33. Morren M, van Dulmen S, Ouwerkerk J, Bensing J. Compliance with momentary pain measurement using electronic diaries: A systematic review. *European Journal of Pain* 2009 Jan;13(4):354–365. doi: 10.1016/j.ejpain.2008.05.010
34. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, Viechtbauer W. The Effects of Sampling Frequency and Questionnaire Length on Perceived Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population. *Assessment* SAGE Publications Inc; 2022 Mar;29(2):136–151. doi: 10.1177/1073191120957102
35. Tsigos C, Kyrou I, Kassi E, Chrousos GP. Stress: Endocrine Physiology and Pathophysiology. In: Feingold KR, Ahmed SF, Anawalt B, Blackman MR, Boyce A, Chrousos G, Corpas E, de Herder WW, Dhatariya K, Dungan K, Hofland J, Kalra S, Kaltsas G, Kapoor N, Koch C, Kopp P, Korbonits M, Kovacs CS, Kuohung W, Laferrère B, Levy M, McGee EA, McLachlan R, Muzumdar R, Purnell J, Rey R, Sahay R, Shah AS, Singer F, Sperling MA, Stratakis CA, Trencle DL, Wilson DP, editors. *Endotext* South Dartmouth (MA): MDText.com, Inc.; 2000. PMID:25905226
36. Russell G, Lightman S. The human stress response. *Nat Rev Endocrinol* Nature Publishing Group; 2019 Sep;15(9):525–534. doi: 10.1038/s41574-019-0228-0

37. Engert V, Vogel S, Efanov SI, Duchesne A, Corbo V, Ali N, Pruessner JC. Investigation into the cross-correlation of salivary cortisol and alpha-amylase responses to psychological stress. *Psychoneuroendocrinology* 2011 Oct 1;36(9):1294–1302. doi: 10.1016/j.psyneuen.2011.02.018
38. Kudielka BM, Gierens A, Hellhammer DH, Wüst S, Schlotz W. Salivary Cortisol in Ambulatory Assessment—Some Dos, Some Don'ts, and Some Open Questions. *Biopsychosocial Science and Medicine* 2012 May;74(4):418. doi: 10.1097/PSY.0b013e31825434c7
39. Kubicek B, Baumgartner V, Prem R, Sonnentag S, Korunka C. Less detachment but more cognitive flexibility? A diary study on outcomes of cognitive demands of flexible work. *International Journal of Stress Management US: Educational Publishing Foundation*; 2022;29(1):75–87. doi: 10.1037/str0000239
40. Christensen M, Treas J. Digital boundary work and work-to-family spillover in Europe: examining the role of digital skills. *Community, Work & Family* Routledge; 2024 Oct;27(5):649–672. doi: 10.1080/13668803.2024.2412700

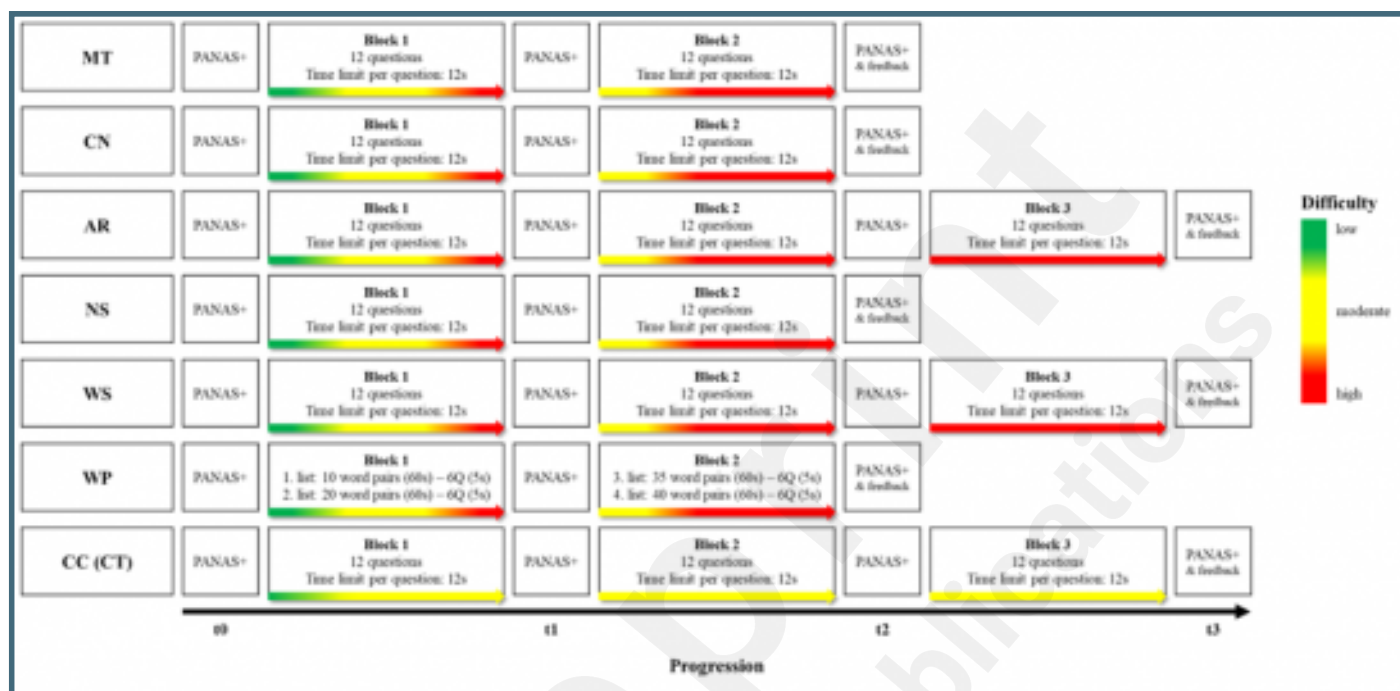
Supplementary Files

Figures

Example screenshots of the Salzburg Mobile Stress Induction (SMSI). SMSI tests in the top line: (a) Matrices, (b) Cube Net, (c) Arithmetic, (d) Number Series. Response feedback in the bottom line: (e) negative feedback, and (f) positive Feedback.



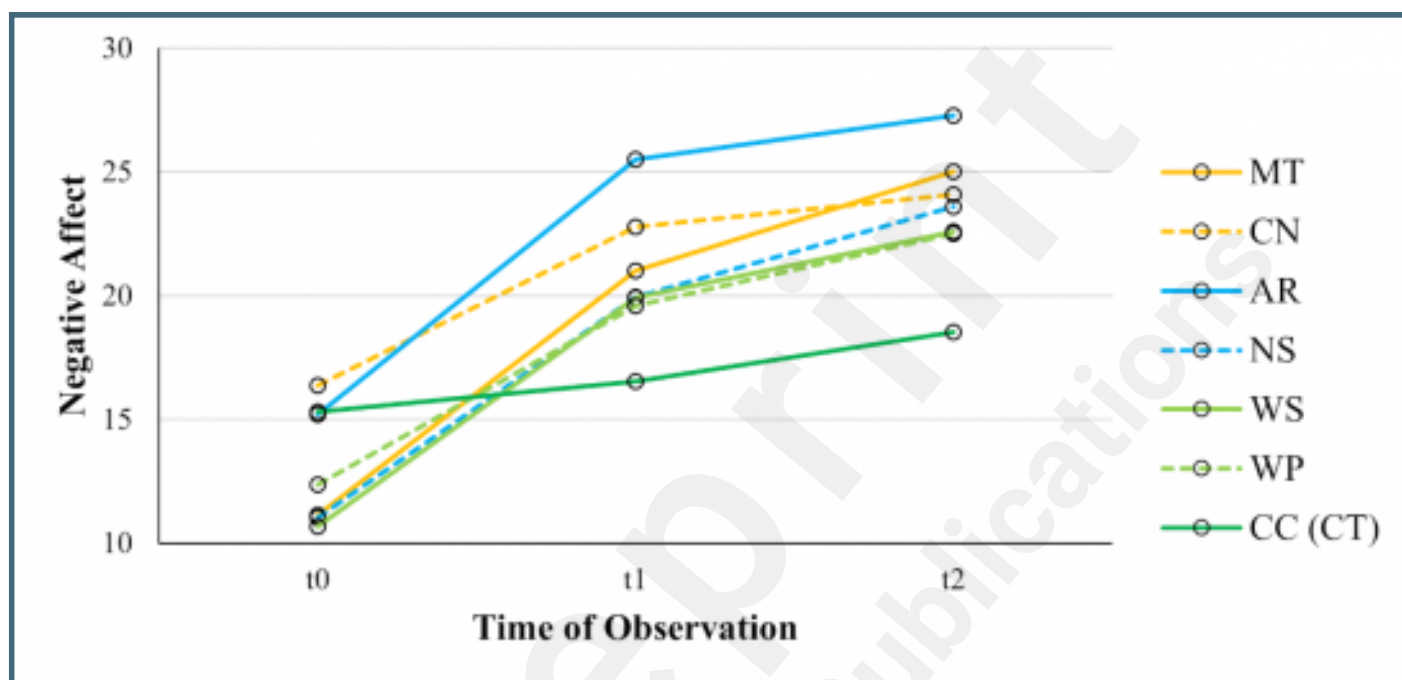
Structure of the Salzburg Mobile Stress Induction tests and assessment procedure. MT=Matrices; CN=Cube Net; AR=Arithmetic; NS=Number Series; WS=Word Scramble; WP=Word Pair; CC (CT)=Caesar Cipher (control test); t0=baseline assessment before the first task block; t1= assessment after the first task block; t2= assessment after the second task block; t3= assessment after the third task block; PANAS+=momentary affective assessment: 10 items of the International Positive and Negative Affect Schedule Short Form [21] plus one single item each on stress and frustration; Feedback=concluding feedback questions on the completed test.



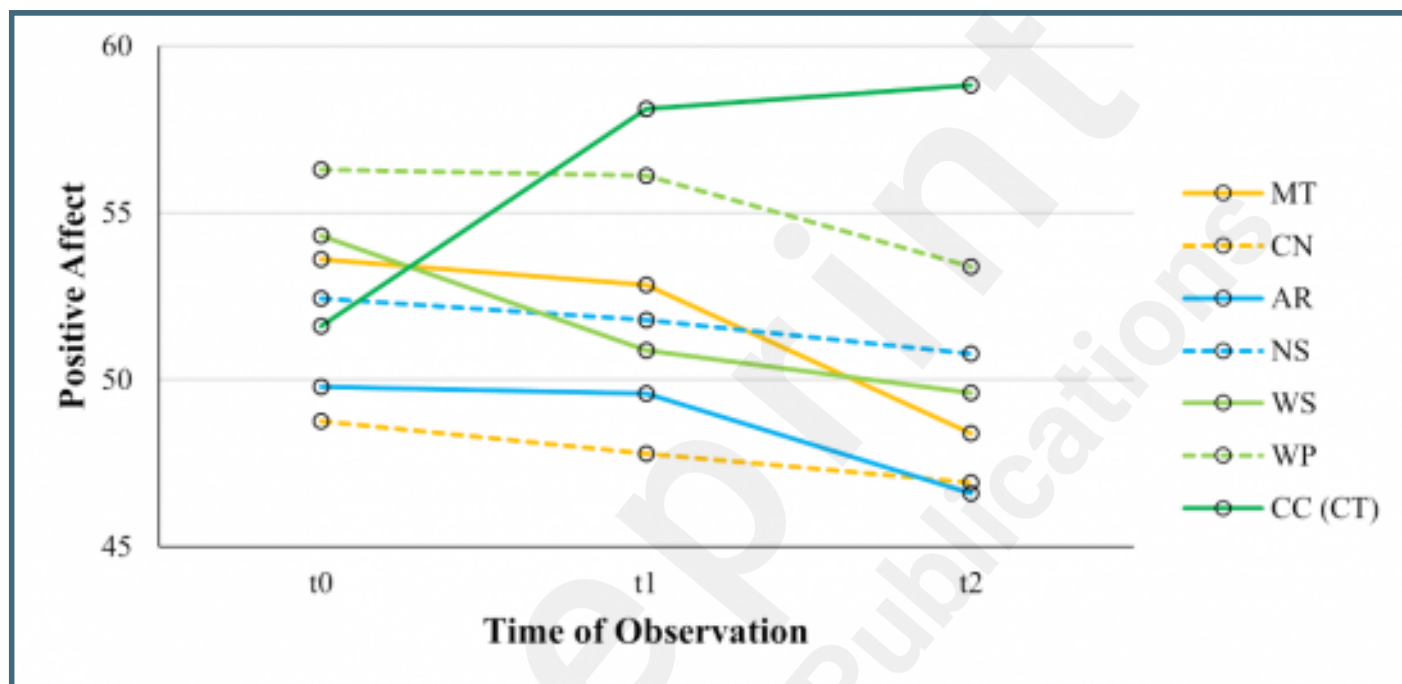
Exemplary schedule of the Salzburg Mobile Stress Induction smartphone study procedure. Colored backgrounds indicate different sections or SMSI test types: Grey=subsidiary parts (welcome message=notification about the pending start of the study; final survey=feedback questions); blue=numeric tests; green=word-based tests (dark green=control test); orange=figural tests. Tests were randomized between participants. NS=Number Series; CC=Caesar Cipher; MT=Matrices; WP=Word Pair; AR=Arithmetic; CN=Cube Net.

	Day 1	Day 2	Day 3	Day 4	Day 5
8 AM					
9 AM	welcome message	NS	CC	MT	WS
10 AM					
11 AM					
12 PM					
1 PM					
2 PM					
3 PM		WP	AR	CN	final survey
4 PM					
5 PM					
6 PM					
7 PM					

Negative affect measured by the Positive and Negative Affect Schedule (PANAS) for each test of the Salzburg Mobile Stress Induction (SMSI) across all participants at baseline (t0), after the first task block (t1) and after the second task block assessments (t2). Orange lines represent the figural tests (Matrices, MT and Cube Net, CN), turquoise lines represent numerical tests (Arithmetic, AR and Number Series, NS), light green lines represent word-based tests (Word Scramble, WT and Word Pair, WP) and the dark green line represents the word-based control test (Caesar Cipher, CC (CT)). Negative affect scores significantly increased from t0 to t1 as well as from t0 to t2 in all tests except the CC control test in which there was only a significant increase from t0 to t2. CN and AR did not significantly increase negative affect from t1 to t2 whereas the other tests did.



Positive affect measured by the Positive and Negative Affect Schedule for each test of the Salzburg Mobile Stress Induction (SMSI) across all participants at baseline (t0), after the first task block (t1) and after the second task block assessments (t2). Orange lines represent the figural tests (Matrices, MT, and Cube Net, CN), turquoise lines represent numerical tests (Arithmetic, AR and Number Series, NS), light green lines represent word-based tests (Word Scramble, WS and Word Pair, WP) and the dark green line represents the word-based control test (Caesar Cipher, CC (CT)). Positive affect scores significantly increase from t0 to t1 as well as from t0 to t2 in the CC control test and significantly decreased from t0 to t1 in WS, from t0 to t2 in MT and NS (trend in AR) as well as from t1 to t2 in the MT, AR and WP. There were no significant changes in positive affect from t0 to t1 in MT, CN, AR, NS and WP, from t0 to t2 in CN, NS and WP, and from t1 to t2 in CN, NS, WS, and CC control test.



Multimedia Appendixes

Example screenshots of the word-based tests of the Salzburg Mobile Stress Induction (SMSI) in the German language used in this study: Task of the Word Scramble test (Figure S1), word pair memory list of the Word Pair test (Figure S2), task of the Word Pair test (Figure S3), and task of the Caesar Cipher control test (Figure S4).

URL: <http://asset.jmir.pub/assets/464e26cdd2058c28e4d91a037775364f.doc>

Video illustrating the complete procedure of the Arithmetic test of the Salzburg Mobile Stress Induction (SMSI). The test was performed in German on a Google Pixel 6a Android (version 15) smartphone in the m-Path App, as used in this study. The video has been modified to 2x playback speed. For proper inspection or reading of specific parts, the video can be paused. Screen recordings were done by the first author in April 2025.

URL: <http://asset.jmir.pub/assets/e04b237fdff1d5b9add486d1d343be41.mp4>

Step-by-step image-based instruction on how to download the m-Path app used in this study on the participants' own smartphone devices as well as on how to register and connect to the research account within the m-Path app. Instructions are in the German language (an English version of the step-by-step image-based instruction as used in the upcoming English version of the Salzburg Mobile Stress Induction is available from the authors).

URL: <http://asset.jmir.pub/assets/48d728f0beb21cf7afa7245b71e91d8c.doc>

Table S1. Table of Cronbach α of the Positive and Negative Affect Schedule (PANAS) subscales negative affect (5 items) and positive affect (5 items) for each test of the Salzburg Mobile Stress Induction and at baseline (t0), after the first task block (t1) and after the second task block assessments (t2; N=100).

URL: <http://asset.jmir.pub/assets/b7416e36ca7e8810daa7f6343bfafef2.doc>

Table S2. Table of the socio-demographic characteristics of the total sample and split between the local university and crowdsourcing subsamples.

URL: <http://asset.jmir.pub/assets/0918fff2bcd9a9e5621d5fc08973ccf9.doc>

Table S3. Table of the mean and SD of the cumulative and average completion times in minutes of the Salzburg Mobile Stress Induction's smartphone procedure and each segment in the total sample and split between the local university and crowdsourcing subsamples. SMSI=Salzburg Mobile Stress Induction.

URL: <http://asset.jmir.pub/assets/cfab216abf6621a237f799e51281e26.doc>

Table S4. Table of the two-way repeated measures ANOVA and contrast statistics for negative affect of the stress-inducing tests (Matrices, Cube Net, Arithmetic, Number Series, Word Scramble, and Word Pair) against the control test (Caesar Cipher) of the Salzburg Mobile Stress Induction (test) from baseline (t0) to after the first task block (t1) and after the second task block assessments (t2; time) controlling for sample (local university, n=69, and crowdsourcing subsample, n=31) as between-subject factor (N=100). Greenhouse-Geisser corrected degrees of freedom were used in the within-subjects effect statistics. η^2 =partial eta squared.

URL: <http://asset.jmir.pub/assets/2bc7c457cd94565b1a9d4953e9799d35.doc>

Table S5. Table of the two-way repeated measures ANOVA and contrast statistics for negative affect of the stress-inducing tests (Matrices, Cube Net, Arithmetic, Number Series, Word Scramble, and Word Pair) against the control test (Caesar Cipher) of the Salzburg Mobile Stress Induction (test) from baseline (t0) to after the first task block (t1) and after the second task block assessments (t2; time) controlling for gender (60/96, women and 34/96, men) as between-subject factor (n=96). Data from 5 participants were missing due to unassignable data from the initial survey to the data from the smartphone study procedure and 1 participant identified as non-binary and was excluded from the analysis. Greenhouse-Geisser corrected degrees of freedom were used in the within-subjects effect statistics. η^2 =partial eta squared.

URL: <http://asset.jmir.pub/assets/98eb0a6c6f9c61a91ca2e35d71888bf3.doc>

Table S6. Table of the two-way repeated measures ANOVA and contrast statistics for positive affect of the stress-inducing tests (Matrices, Cube Net, Arithmetic, Number Series, Word Scramble, and Word Pair) against the control test (Caesar Cipher) of the Salzburg Mobile Stress Induction (test) from baseline (t0) to after the first task block (t1) and after the second task block assessments (t2; time) controlling for sample (local university, n=69, and crowdsourcing subsample, n=31) as between-subject factor (N=100). Greenhouse-Geisser corrected degrees of freedom were used in the within-subjects effect statistics. η^2 =partial eta squared.

URL: <http://asset.jmir.pub/assets/bdac494f439f83df6dcd435cfa539413.doc>

Table S7. Table of the two-way repeated measures ANOVA and contrast statistics for positive affect of the stress-inducing tests (Matrices, Cube Net, Arithmetic, Number Series, Word Scramble, and Word Pair) against the control test (Caesar Cipher) of the

Salzburg Mobile Stress Induction (test) from baseline (t0) to after the first task block (t1) and after the second task block assessments (t2; time) controlling for gender (60/100, women and 34/100, men) as between-subject factor (N=100). Data from 5 participants were missing due to unassignable data from the initial survey to the data from the smartphone study procedure and 1 participant identified as non-binary and was excluded from the analysis. Greenhouse-Geisser corrected degrees of freedom were used in the within-subjects effect statistics. η^2 =partial eta squared.

URL: <http://asset.jmir.pub/assets/58737bb8001df0bfc5a347c8a15f7d24.doc>

