Psychoneuroendocrinology 110 (2019) 104437

Contents lists available at ScienceDirect

Psychoneuroendocrinology

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

A meta-analysis of cortisol reactivity to the Trier Social Stress Test in virtual environments

Emily C Helminen[a], Melissa L Morton[a], Qiu Wang[b], Joshua C Felver[a,*]

[a] Department of Psychology, 430 Huntington Hall, Syracuse University, Syracuse, NY, 13244, United States
[b] Department of Higher Education, 350 Huntington Hall, Syracuse University, Syracuse, NY, 13244, United States

ARTICLE INFO

Keywords:
Cortisol
Stress
TSST
Virtual reality
Stress reactivity
Psychosocial stress

ABSTRACT

Background: Maladaptive responses to stressors can lead to poor physical and psychological health outcomes. Laboratory studies of stress induction commonly use the Trier Social Stress Test (TSST). The TSST has been shown to reliably induce a stress response, most commonly measured *via* cortisol reactivity. Recently, researchers have used virtual environment versions of the TSST (V-TSST) in place of the traditional TSST. The V-TSST has many advantages over the traditional TSST, including increased standardization and use of fewer resources, but V-TSST has yet to be quantitatively reviewed and compared to the traditional TSST. This review aims to quantifying the effectiveness of V-TSST with a meta-analysis of cortisol response effects and identify potential moderating variables that are more likely to induce a cortisol response with V-TSST.

Methods: Literature searches were conducted including the key words Trier Social Stress Test, TSST, and virtual reality. Thirteen studies were included in this meta-analysis after meeting the inclusion criteria of utilizing a V-TSST and having cortisol measurements at baseline and peak stress to assess cortisol reactivity. The standardized mean gain effect size was used.

Results and discussion: There was a medium average effect size (ES_{sg} = 0.65) across all studies for increase in cortisol from baseline to peak measurement. Significant moderating effects were seen for participant age, sex, and level of immersivity of the virtual environment. Studies in which participants were under 25 years old, or all male, showed greater effect sizes for cortisol reactivity. Virtual environments that were more immersive also evidenced greater effect sizes. Although the V-TSST is effective at inducing psychosocial stress, the magnitude of this response is less than the traditional TSST. Based on these results, recommendations for future research are provided.

1. Introduction

Stress is a ubiquitous human experience, and although the human body's response to stressful events is evolutionarily adaptive, some individual's patterns of stress responding are maladaptive (McEwen, 2007). Contemporary stressors are often social-evaluative in nature and chronic maladaptive responses to these stressors can lead to poor physical and psychological health outcomes (Schneiderman et al., 2005; Dickerson et al., 2004). Thus, understanding the nuances by which individuals respond to stress is of critical importance for improving health outcomes, and replicable and standardized methodologies for studying stress responding is necessary for future high-quality research.

To study stress in controlled conditions, reliable methods for inducing a stress response are needed. Among the various methods for measuring stress reactions, physiological stress responses provide a direct relation to health outcomes and objectivity in measurement (Chrousos, 2009; Dickerson et al., 2004). The main stress response system in the body is the hypothalamic-pituitary-adrenal (HPA) axis. One measure of the HPA axis, cortisol, is particularly valuable and appealing given its reliability as a marker of sympathetic activation, and it has a robust literature base relative to other physiological measures of stress response (Allen et al., 2014; Dickerson and Kemeny, 2004). Upon activation, the HPA axis sets off a cascade of physiological reactions that culminates in the release of cortisol by the adrenal glands. Research consistently indicates that cortisol release peaks at approximately 20–30 minutes after the onset of a stressor (Engert et al., 2013; Foley and Kirschbaum, 2010), providing a reliable time window for biological sampling of the stress response.

In an effort to study stress reactivity in a laboratory setting, the Trier

[*] Corresponding author.
E-mail address: jcfelver@syr.edu (J.C. Felver).

https://doi.org/10.1016/j.psyneuen.2019.104437
Received 6 February 2019; Received in revised form 12 August 2019; Accepted 5 September 2019

Social Stress Test (TSST) was developed to induce a physiological stress response (Clemens Kirschbaum et al., 1993). In the original protocol, Kirschbaum et al. (1993) detailed a three-part task that included a speech preparation portion, a speech delivery portion, and an arithmetic portion. Participants are brought into a room with a one-way mirror and seated before two research assistants posing as confederates in white lab coats. They are told to prepare for a speech in which they discuss why they are the perfect candidate for their ideal job. Participants then complete the speech portion and immediately after are told to count backwards from 1022 by 13. Throughout the entire experiment, confederates are instructed to maintain neutral affect, regardless of what the participant says or does (Birkett, 2011; Kirschbaum et al., 1993). The original protocol effectively induces stress *via* components of social evaluation and the unpredictability of the confederates response to the subject's performance (Dickerson and Kemeny, 2004). Dickerson and Kemeny (2004) argue that the combination of public speaking and cognitive components create high-levels of social-evaluative threat and unpredictability, and that these elements in turn lead to a reliable stress response as measured by measurement of cortisol post-TSST onset.

Despite the TSST being the most widely used method for acute psychosocial stress induction (Allen et al., 2016), there are methodological limitations with this technique. Firstly, the traditional TSST is resource intensive. Based on the original protocol (Clemens Kirschbaum et al., 1993), the traditional TSST requires at least three research assistants to run each participant (*i.e.*, one to greet the participant and give instructions, and two to serve as confederates). Additionally, it requires access to a room outfitted with a one-way mirror to further create the experience of being evaluated. Secondly, although many aspects are highly standardized, the confederates themselves as individuals are often variable, as laboratories running a traditional TSST study for any length of time will often need to employ many research assistants to serve as confederates in the experiment. Confederate age, sex, race, and ability to maintain a neutral affect may vary across individuals, and the efficacy of the traditional TSST could be influenced as a result of these variations. For example, a recent meta-analysis shows that both changes in affective display and sex composition of the confederates are significantly associated with reduced cortisol response to the traditional TSST (Goodman et al., 2017). Thus, devising a TSST protocol that reduces the resource intensiveness and variability of the procedure while maintaining the social-evaluative and unpredictable nature of the task would remove barriers to conducting high-quality stress research.

1.1. Virtual environments in stress research

One way to address the resource intensive and variable aspects of the traditional TSST protocol is to embrace emerging technologies, including virtual environments (VEs). Stanković (2015) broadly defines virtual environments as "artificial spaces in some way separated from the physical world" (p. 9). There are many different types of VEs, and in this review, screen-based environments are included in this definition of virtual environments. Due to the availability of screen-based communication technology (*e.g.*, Skype and Facetime), together with the development and increased adoption of commercially available immersive technologies (*e.g.*, Google Cardboard, GearVR, Oculus), virtual environments are now widely accessible and commonly used by many people.

Virtual environments have been widely embraced in exposure research and are successful at inducing fear (Diemer et al., 2014) and treating a variety of anxiety disorders (Powers and Emmelkamp, 2008). Similarities exist between exposure therapy for anxiety disorders and implementing the traditional TSST, with both protocols intending to create a stressful experience for participants. Traditional (*i.e.*, *in vivo*; Bouchard et al., 2017) exposure therapy also shares similar barriers to implementation with the traditional TSST. It has the potential to be resource-intensive, often requiring access to locations or materials that will successfully induce fear (*e.g.*, a tall building to work with a fear of heights). When exposure therapy is implemented in virtual environments, it is less resource-intensive and more cost-effective (Emmelkamp, 2005). Traditional exposure therapy can also be widely variable in implementation, and it is also difficult to implement intensive exposure in some specific cases, such as airplane flying phobias without access to a flying airplane (Krijn et al., 2004). Given that virtual environments have successfully been used to induce fear and increase standardization in exposure therapy research, other domains of scientific inquiry may similarly benefit from adopting this technology to reduce barriers to conducting high-quality research.

In recent years, scientists have begun to employ virtual environments to induce acute stress in human research participants. To reduce the labor-intensiveness and increase replicable standardization of the TSST, virtual environment TSST protocols (V-TSST) have been developed, such as by using 3D projections (Jönsson et al., 2010), immersive headgear (Shiban et al., 2016; Kelly, Matheson et al., 2007), and screen-based displays (Fallon, Careaga et al., 2016; Hawn, Paul et al., 2015;Santos-Ruiz et al., 2010). Although these different V-TSST have all been successful in inducing acute stress in human participants, to-date there has not been any critical analyses of the entirety of these varying protocols.

1.2. Cortisol reactivity to the TSST

The traditional TSST protocol has been shown to successfully induce a cortisol response in hundreds of studies since it was developed. In a recent meta-analysis, Goodman and colleagues (2017) provided evidence that across studies (n = 186), the traditional TSST reliably induced strong cortisol responses (d = 0.93). However, some protocol variations, such as confederate affect and sex, diminished participant cortisol responding (Goodman et al., 2017). In recent years, promising research employing various V-TSST protocols has also provided evidence for a consistent cortisol stress response; however, this emerging body of V-TSST research has yet to be systematically and meta-analytically evaluated and compared to the original TSST stress response outcomes.

Although V-TSST offers the advantages of fewer necessary personnel resources and the ability to standardize aspects of the experiment (*e.g.*, composition of the confederates, physical setting), in order for the field of stress research to consider the adoption of the virtual versions of the TSST, it is critical to demonstrate that V-TSST elicits a reliable and significant physiological stress response (*e.g.*, cortisol response) that is similar to the traditional TSST protocol. If the V-TSST can produce a similar stress response, laboratory-based studies of stress may be more inclined to adopt this new technology. In summary, employing the V-TSST offers the notable advantages of being less resource-intensive and more standardized than the traditional TSST, and with systematic adoption of the V-TSST, both the quantity and quality of laboratory-based stress studies may increase, leading to a better understanding of the stress response and ultimately providing a useful tool for those interested in study human physiological stress responding.

1.3. Study aims

A virtual version of the TSST could provide a higher level of standardization in evaluating the effects of acute psychosocial stress induction by decreasing the resource-intensiveness and variability due to confederate characteristics. However, the effectiveness of V-TSST on stress induction has yet to be systematically or quantitatively evaluated. In this paper, we aim to critically evaluate the literature on the effectiveness of V-TSST by (a) quantifying the effectiveness of V-TSST with a meta-analysis of cortisol response effects, (b) conducting moderator analyses to determine if V-TSST study characteristics are related to obtained effect sizes, and (c) providing recommendations for future V-

TSST based on findings obtained.

2. Methods

2.1. Search strategy

A search of online databases for studies using any type of V-TSST was performed with the keywords: "virtual reality" and ("TSST" or "Trier Social Stress Test"). The databases included: PubMed, PsycInfo, PsycArticles, and MEDLINE. All English-language articles were collected up until November 2018. This initial search yielded 50 potential publications. Reference lists from these papers were then reviewed to identify additional publications that were not captured in the initial database searches, yielding a sample size of 55 publications.

2.2. Inclusion/exclusion criteria and study selection

For inclusion in this review, studies were required to be conducted with adults (age > 18 years), use a type of V-TSST, and collect cortisol samples both at baseline and peak of acute stress reactivity. The peak of stress varied between studies, with a range of 18–30 minutes post-stressor onset ($M = 24.5$, $SD = 4.74$). The V-TSST was defined as any type of TSST that included participants being exposed to both a speech task and an arithmetic task where either the confederates or the participant was in an environment other than where the participant was physically located (*i.e.*, a virtual environment). This includes immersion into a virtual world where participants interact with study confederates through the use of 3D projections, immersive headsets, or any digital screens.

Studies were excluded if they were conducted with youth (age < 18 years), as different age groups have been shown to have different stress reactivity responses to acute psychosocial stress (Allen et al., 2016). Studies were also excluded if they removed a defining component of the TSST (*i.e.*, the speech task or the math task) that was in the original protocol established by Kirschbaum et al. (1993). This is due to the fact that the TSST protocol was designed to be maximally stressful for the participants, and research validating the cortisol reactivity to the TSST indicated the greatest effect sizes for studies that included both components of the TSST (Dickerson and Kemeny, 2004).

After duplicates were removed ($n = 28$), the remaining articles' abstracts were reviewed to determine if they met the aforementioned selection criteria, resulting in 5 studies being removed. After this screening, full texts of the remaining were analyzed ($n = 22$). Inclusion and exclusion criteria were applied, and further studies were identified for removal due to not using a V-TSST as defined previously ($n = 3$), not using both speech and math tasks ($n = 3$), lacking cortisol measurements ($n = 2$), and inability to get cortisol data from the paper or corresponding author ($n = 1$). The remaining studies ($n = 13$) were included as a part of the meta-analysis. Fig. 1 details the flowchart of selection of studies into this meta-analysis.

2.3. Data coding and extraction

For each of the papers included in this review, data on race, sex, sample size, and type of virtual environment (*e.g.*, immersive headsets, flat screen projection) were extracted. For studies that were evaluating both clinical (*e.g.*, adults with exhaustion disorder) and non-clinical healthy populations (Linninge et al., 2018; Fallon, Careaga et al., 2016; Jönsson et al., 2015), only the healthy population values were extracted so as to not add any confounding variables when looking at the effectiveness of V-TSST in eliciting a stress response, as various clinical populations have differing stress reactivity in response to the TSST (Zänkert et al., 2018; Allen et al., 2014). In studies where groups were separated by non-clinical characteristics (*e.g.*, performance on the Iowa Gambling Task; Santos-Ruiz et al., 2012) data for groups were combined (Higgins and Green, 2011). Details of the selected study details are presented in Table 1.

The first and second authors independently performed the data coding and achieved an inter-rater reliability (IRR) score of 96.9% (calculated as # of agreements / total # of coded data points). Any disputes between coders were discussed until a consensus was reached.

When reported, means and standard deviations for cortisol values at baseline and height of stress were extracted. For studies that did not explicitly report these values ($n = 10$), two independent coders (*i.e.*, the first author and a trained research assistant) used the plot extraction software WebPlotDigitizer (Rohatgi, 2018) to obtain precise values from cortisol graphs. These methods and this software have been previously shown to be accurate at determining plot values, in addition to allowing coders to achieve a high IRR score and Pearson's r correlation (Drevon et al., 2017). For extraction, mean cortisol values were graphed and error bars represented the standard error of the mean (SEM). Once values were determined, the standard deviations were calculated from the SEM *via* the following formula:

$$SD = SEM * \sqrt{n-1}$$

Two separate coders performed the cortisol data extraction independently and achieved an IRR score of 95.7 percent (calculated as # of agreements / # of data points). The Pearson's r correlation between coders was $r = 0.9999$, $p < 0.001$. In accordance with Drevon et al. (2017), coders were considered to be in agreement if the extracted cortisol values from each were within 1% of each other, relative to the range of the y-axis. Discrepancies between coders were discussed and both coders re-extracted the discrepant data points. This was sufficient to reach consensus on all data points.

2.4. Publication bias

Publication bias was assessed using Duval and Tweedie's (2000) recommendations. In this method, individual study effect sizes and standard errors are plotted and visually inspected to see if they form a funnel shape around the overall mean effect size. Often, published studies show a bias towards larger effect sizes. With the trim and fill method, potentially missing studies are included in the funnel plot and a new overall mean effect size is calculated to see how much the missing studies negatively affect the original overall mean effect size calculation.

Additionally, the fail-safe N (Orwin, 1983; Rosenthal, 1979) was calculated. This computation estimates the number of unpublished studies that would need to show insignificant results for the overall effect size to be reduced to a specified level. The greater the fail-safe N, the more confidence there is in the meta-analytic results.

2.5. Data analysis

Since most of the studies in this review did not compare V-TSST to traditional TSST, within-subjects effect sizes were used to determine the effectiveness of V-TSST in eliciting a cortisol response. Following recommendations for calculating for pre-post contrasts detailed by Lipsey and Wilson (2001), the standardized mean gain effect size statistic (ES_{sg}) was calculated for each study. This calculation includes a pooled standard deviation term (SD_p). Due to the nature of the SD_p calculation (see Lipsey and Wilson, 2001, p. 44), this effect size is similar to a Cohen's d effect size as is interpreted the same way (see Goodman et al., 2017). However, to be consistent throughout the document, we will refer to effect sizes as ES_{sg}. The standardized, rather than unstandardized, mean gain values were selected due to the varying metrics of report cortisol values (*e.g.* log cortisol, nmol/L). The following formula was used:

$$ES_{sg} = \frac{M_{peak} - M_{base}}{SD_p}$$

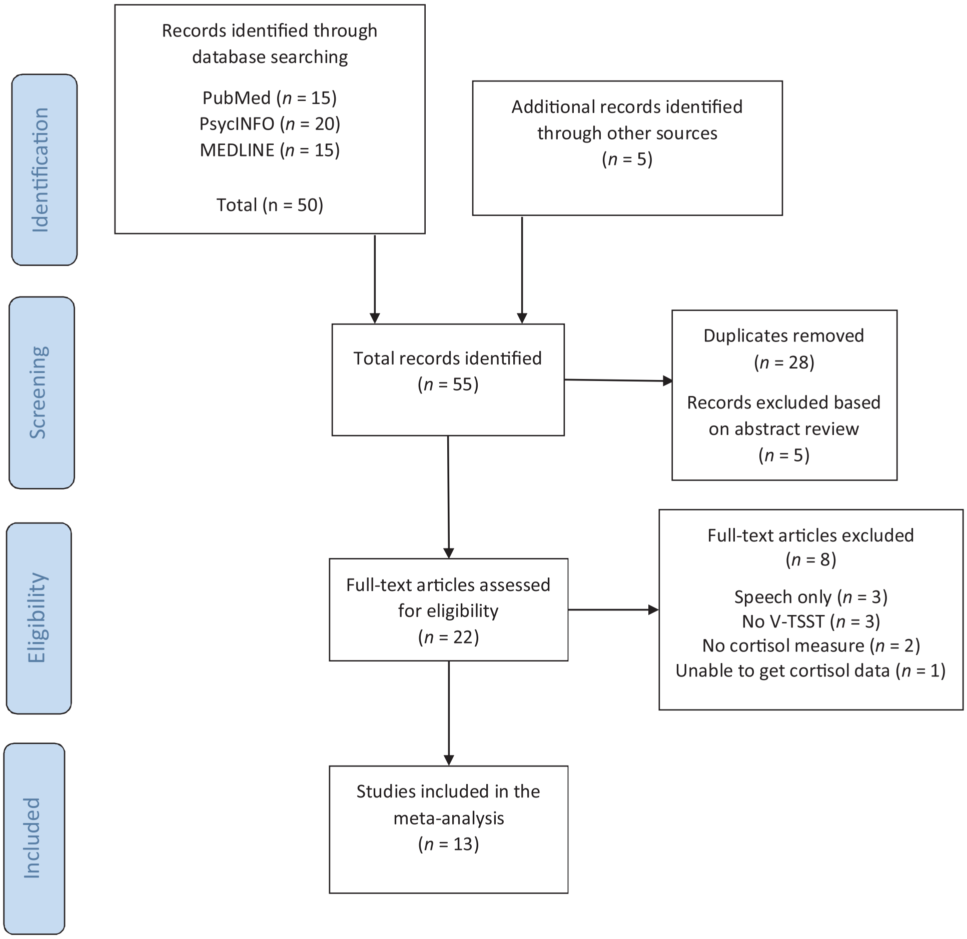
Identification
Screening
Eligibility
Included
Records identified through database searching
PubMed (n = 15)
PsycINFO (n = 20)
MEDLINE (n = 15)
Total (n = 50)
Additional records identified through other sources (n = 5)
Total records identified (n = 55)
Duplicates removed (n = 28)
Records excluded based on abstract review (n = 5)
Full-text articles assessed for eligibility (n = 22)
Full-text articles excluded (n = 8)
Speech only (n = 3)
No V-TSST (n = 3)
No cortisol measure (n = 2)
Unable to get cortisol data (n = 1)
Studies included in the meta-analysis (n = 13)

Fig. 1. Preferred reporting items for systematic reviews and meta-analysis (PRISMA) flowchart.

Table 1
Study characteristics.

Study	Sample Size	Participant Sex Composition (Male/Female)	Mean Age (Years)	Confederate Type	Virtual Environment Type	Immersive (Yes/No)
(Annerstedt 2013)	10	M: 100% F: 0%	28.1	Avatars	CAVE	Yes
(Fallon 2016)	16	M: 50% 50%	19.31	Avatars	3D Screen	No
(Fich 2014)	14	M: 100% F: 0%	23.9	Avatars	CAVE	Yes
(Hawn 2015)	21	M: 54.5% F: 45.5%	38.45	Pre-recorded video of real people	2D Screen	No
(Jönsson et al., 2010)	10	M: 100% F: 0%	28.3	Avatars	CAVE	Yes
(Jönsson et al., 2015)	20	M: 50% F: 50%	49.2	Avatars	CAVE	Yes
(Kelly 2007)	46	M: 49.6% F: 50.4%	21.45	Pre-recorded video of real people	HMD	Yes
(Linninge 2018)	23	M: 100% F: 0%	24.7	Avatars	CAVE	Yes
(Montero-López et al., 2018)	18	M: 0% F: 100%	33.17	Avatars	3D Screen	No
(Santos-Ruiz 2010)	21	M: 28.6% F: 71.4%	24	Avatars	3D Screen	No
(Santos-Ruiz 2012)	38	M: 0% F: 100%	28	Avatars	3D Screen	No
(Shiban 2016)	5	M: 100% F: 0%	23.76	Avatars	HMD	Yes
(Zimmer 2019)	24	M: 100% F: 0%	24.93	Avatars	HMD	Yes

Notes: CAVE = Cave Automatic Virtual Environment; HMD = head mounted display.

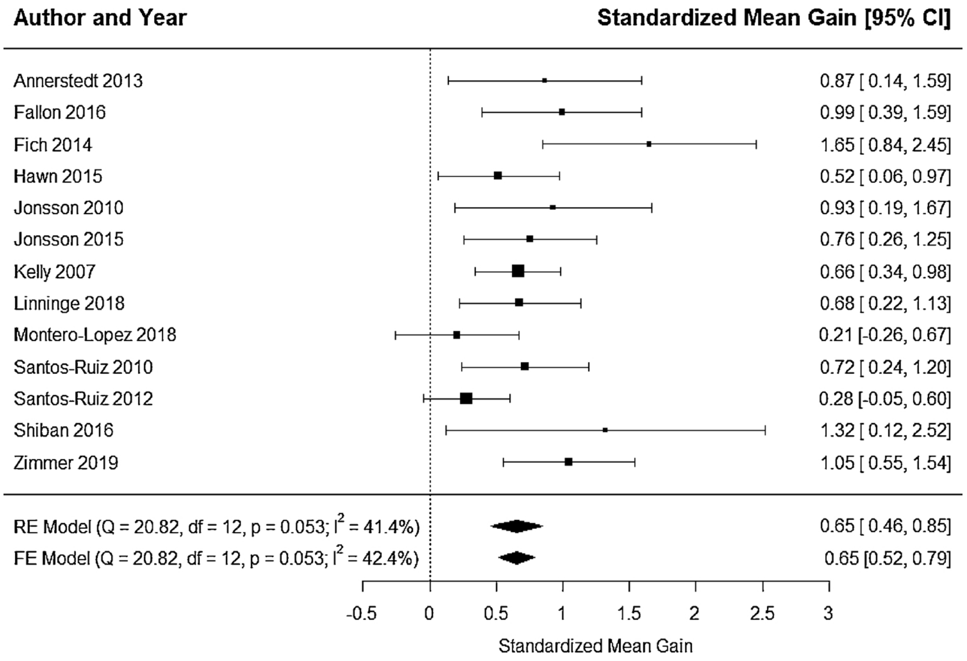

Fig. 2. Forest plot of individual study effect sizes and overall effect sizes with both a random-effects (RE) model and a fixed-effects (FE) model.

where M_{peak} = mean cortisol value at peak stress

M_{base} = mean cortisol value at baseline

SD_p = pooled standard deviation of cortisol values at peak and baseline

With the standardized mean gain effect size, pre-post r correlation coefficients are needed to compute standard error for the effect size. To compute the pre-post r correlation, access to individual subject data from each study would be necessary. Since that was not possible, a sensitivity analysis was implemented, with r values ranging from 0.1 to 0.9 to determine if changes to this correlation coefficient would have a significant effect on computing the overall mean effect size.

Analysis of homogeneity is often used to detect whether there are between-study differences that go beyond sampling error. Homogeneity analysis was performed on the data for this review using the Q statistic, as recommended by Lipsey and Wilson (2001).To further elucidate the effectiveness of V-TSSTs, moderation analyses were also conducted on the effects of sex, age, and virtual environment immersivity in the open-source statistical software R and Metafor (Viechtbauer, 2017).

Different virtual environments were determined to be immersive or non-immersive based on how completely inputs from the real world were replaced with inputs from the virtual environment. Virtual environments that replaced the all visual and audio cues from the real world with virtual visual and audio cues were considered to be immersive. For example, both 3D projections covering the entire visual space in front of the participant and head mounted devices (HMDs) paired with headphones would be considered immersive. Screen-based environments (*e.g.*, computer or television) are not considered immersive in this review because peripheral visual cues are still coming from the real world. Of the studies, five used 3D projection technology (*i.e.*, Cave Automatic Virtual Environment; CAVE). CAVE systems and other similar 3D systems rear-project the virtual environment onto three walls in front of the participant and are paired with head tracking and stereoscopy equipment to achieve the 3D effect. Of the remaining studies, five used screen-based technology, including both 3D screens (n = 3) and 2D screens (n = 2). Three studies used immersive head-mounted displays (HMDs). The CAVE systems and HMDs were considered to be immersive, whereas the 2D and 3D screens were considered to be non-immersive (Bowman and McMahan, 2007). Table 1 summarizes the type of virtual environment used by each study.

Age and sex were initially analyzed as continuous variables using modified weighted least squares regression (Lipsey and Wilson, 2001). In this method, each effect size is weighted by the inverse of its variance and two indices (*i.e.*, Q_M and Q_E) are calculated to determine overall fit of the model. If Q_M is significant, this indicates that the regression model accounts for significant variability in the effect sizes. If significant, the Q_E statistic indicates that there is residual variability in effect sizes that is not accounted for in the regression model.

Age, sex, and immersivity were further analyzed using the meta-analytic analog to the ANOVA. With this method, effect sizes are grouped into mutually exclusive categories (*e.g.*, immersive or non-immersive) and homogeneity tests are calculated for variability within each group (Q_W) and between the groups (Q_B). Because it uses homogeneity statistics, the meta-analytic analog to the ANOVA takes into account the weight given to each effect size. The Q_W statistic can be calculated for each subgroup, and if Q_W is significant for a subgroup, this indicates that there is still significant variability within that subgroup. If Q_B is significant, this indicates that the categorical separation variable accounts for variability between each group of effect sizes.

3. Results

A total of 13 studies were included for the meta-analysis. Eight studies were considered immersive, and five studies were coded as non-immersive based on their V-TSST protocol description. For the panel of confederates, most studies (n = 11) in this review used avatars, or virtual persons, and two studies used pre-recorded video of real confederates. Six studies included only male participants, five included a mixture of male and female participants, and two studies included only female participants. The mean age of participants for each study ranged from 19.3 to 49.2 years old (M = 28.25, SD = 8.03). Overall, 61.5% of studies were immersive, average sex composition was 64.1% male (SD = 38.63). Details on each study are presented in Table 1.

3.1. Cortisol reactivity to V-TSSTs

Individual study and average study effect sizes are presented in Fig. 2. Initially, effect sizes were analyzed with a random-effect model. Test of heterogeneity was non-significant (Q = 20.82, df = 12, p = 0.053), indicating that the variance between studies was not shown to be greater than what may occur due to sampling error (Lipsey and

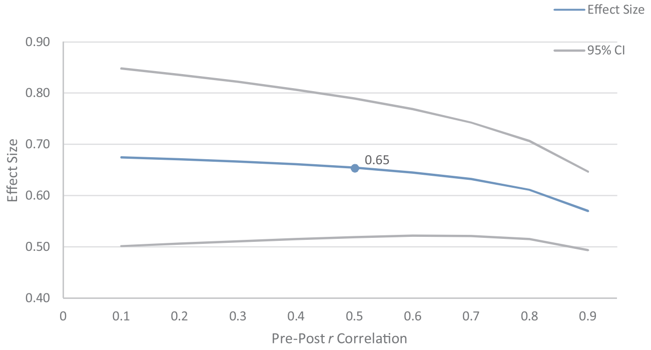

Fig. 3. Sensitivity analysis of pre-post r correlation from $r = 0.1$ to $r = 0.9$.

Wilson, 2001). Due to these results, effect sizes were re-analyzed with a fixed-effect model that assumes non-significant heterogeneity between studies. Both sets of results are included in Fig. 2.

The average effect size across all studies for increase in cortisol from baseline to peak measurement) was $ES_{sg} = 0.65$ (SE = 0.069), which is considered a medium effect based on Cohen (1988) recommendations. For individual studies, effect size calculations for cortisol reactivity ranged from small to large (ES_{sg} range = 0.21–1.65).

3.1.1. Sensitivity analysis for pre-post r correlation

Calculation of the standard error for the standardized mean gain effect size requires the pre-post r correlation coefficient, which subsequently affects the overall mean effect size calculation. Since we only had access to aggregated data and not the individual correlation values, we could not calculate mean pre-post correlations for each study as they were not reported. To explore whether the results obtained may have been sensitive to the correlation value between pre- and post-values, effect sizes were calculated first with an estimate of $r = 0.5$, and then a sensitivity analysis was conducted for r values ranging from low (0.1) to high (0.9) correlations. Fig. 3 details the overall mean effect size with changing pre-post r correlation coefficients, along with their respective confidence intervals. Resulting effect sizes ranged from 0.57 to 0.67 across r values. As all effect size values obtained in this sensitivity analysis are all considered medium effect sizes falling within a relatively narrow range of values, the assumption of moderate correlation value of $r = 0.5$ does not seem significantly alter results obtained.

3.2. Moderators of cortisol reactivity to V-TSST

Moderation analyses were conducted looking at effects of sex, age, and immersivity of virtual environment. For each analysis, studies were either looked at as continuous variables with weighted meta-regression analyses, or categorized into two different groups and evaluated with the analog to analysis of variance (ANOVA) detailed in Lipsey and Wilson (2001). For the categorical analyses, the number of categories was limited to two due to the small number of V-TSST studies ($n = 13$). Fig. 4 plots the mean group effect sizes from these moderator analyses.

3.2.1. Sex and cortisol reactivity

Sex was first analyzed as a continuous variable as percentage of male participants. Weighted regression analyses using percent male participants as the predictor variable was performed on the study data. The moderation effect was significant ($Q_M = 13.02$, $df = 1$, $p < 0.001$), while the residual effect (*i.e.*, the heterogeneity of the data after accounting for the moderation analysis) was non-significant ($Q_E = 7.79$, $df = 11$, $p = 0.732$).

To further display the effects of sex on cortisol reactivity, studies were split into two groups, including a group with studies including only male participants ($n = 6$), and a group of studies including either participants with both sexes ($n = 5$) or only female participants ($n = 2$). There were no significant differences within each group ($Q_W = 13.34$, $df = 11$, $p = 0.27$). However, between group differences were significant ($Q_B = 7.48$, $df = 1$, $p = 0.006$), with groups with male only participants ($M = 0.96$, $SE = 0.09$) having a greater mean effect size than groups with mixed or all female participants ($M = 0.54$, $SE = 0.08$).

3.2.2. Age and cortisol reactivity

Age was first analyzed as a continuous variable. Weighted regression analyses using mean age of participants as the predictor variable was performed on the study data. Neither the moderation effect ($Q_M = 1.19$, $df = 1$, $p = 0.274$) nor the residual effect were statistically significant ($Q_E = 19.62$, $df = 11$, $p = 0.051$).

To further consider the effects of age on V-TSST effectiveness, studies were separated into two groups, one with mean ages under 25 years old and one with mean ages over 25 years old. The age of 25 was chosen as a cutoff because it is considered the age at which the brain has fully matured (Arain et al., 2013). When the groups were split up, there were no significant differences within each group ($Q_W = 13.82$, $df = 11$, $p = 0.24$). However, between group differences were significant ($Q_B = 7.00$, $df = 1$, $p = 0.008$), which indicated that the effect size for the under 25 group ($M = 0.83$, $SE = 0.10$) was significantly greater than the overall mean effect size of the over 25 group ($M = 0.47$, $SE = 0.10$).

3.2.3. Immersivity and cortisol reactivity

Studies were categorized as being immersive ($n = 8$) or non-immersive ($n = 5$) according to our previous definition, and differences between these groups were assessed. When the groups were split up, there were no significant differences within each group ($Q_W = 13.54$, $df = 11$, $p = .26$). However, between group differences were significant ($Q_B = 7.28$, $df = 1$, $p = 0.007$), with the mean effect size for immersive environments ($M = 0.83$, $SE = 0.07$) being greater than the mean effect

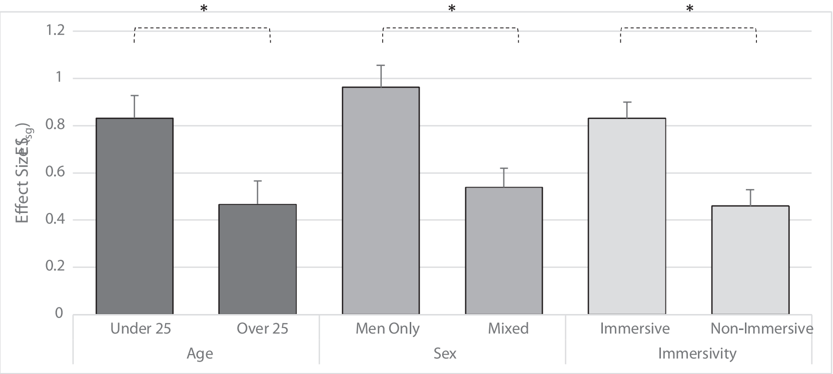

Fig. 4. Effect sizes for moderator analyses of age, sex, and immersivity. * indicates p < 0.01.

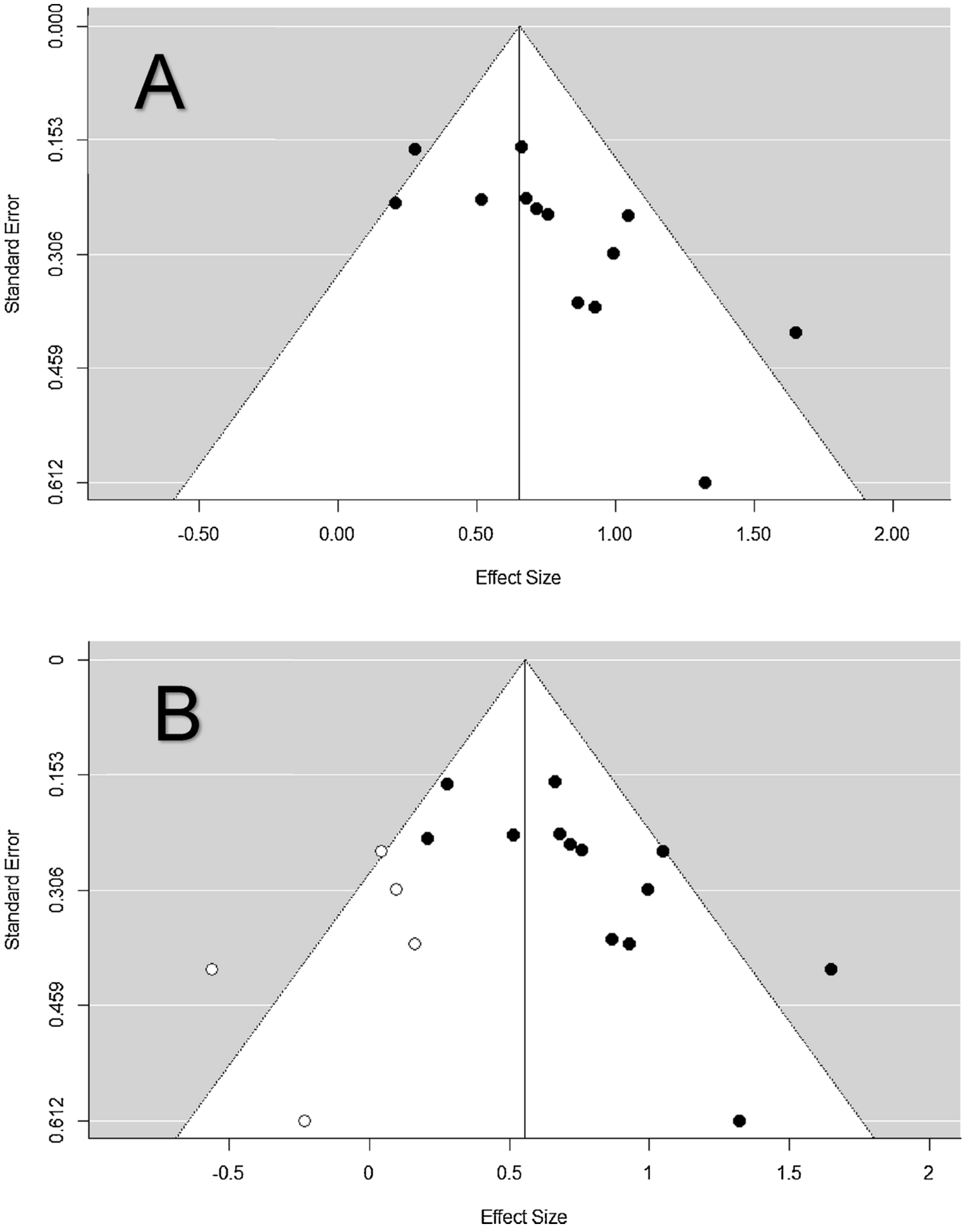

Fig. 5. Funnel plot of studies included in this meta-analysis (A), and funnel plot including potentially missing studies (open circles) using the trim and fill method (B).

size of non-immersive environments (M = 0.46, SE = 0.07). A chart of mean effect sizes for each moderator (Fig. 4) shows these differences more distinctly.

3.3. Examination of publication bias

Publication bias was examined with funnel plots of effect size and standard error (see Fig. 5). When effect sizes for each study were graphed against their respective standard errors, the funnel plot showed a skew towards larger effect sizes when there was greater standard error. To account for publication bias, the trim and fill method by Duval and Tweedie (2000) was performed on the data, resulting in several possible missing studies (n = 5) on the left side (*i.e.*, lower effect sizes) of the funnel plot. Analysis with the missing studies resulted in a slightly lower overall effect size (ES_{sg} = 0.56, SE = 0.13), suggesting that there may be a slight publication bias towards a larger effect size, but that the overall medium effect size is robust to addition of studies with smaller effect sizes.

Further confidence in the meta-analytic results of this review can be established by computing the fail-safe N (Orwin, 1983; Rosenthal, 1979). This computation estimates the number of unpublished studies that would need to show insignificant results for the overall effect size to be reduced to a specified level (*e.g.*, dropping from 0.65 to 0.2). For this meta-analysis, to reduce the overall effect size of ES_{sg} = 0.65 to a small effect size of ES_{sg} = 0.2 or less, there would have to be at least 29 unpublished studies that show null results. Given that our sample size is only 13 studies, and considering that the field of V-TSST research is still relatively new, it is highly unlikely that there are that many unpublished V-TSST studies showing non-significant results.

4. Discussion

The primary goal of this meta-analysis was to evaluate and quantify the effectiveness of V-TSST in eliciting a cortisol response. After conducting a meta-analysis of 13 V-TSST studies examining cortisol reactivity, there is evidence that V-TSST is effective in producing a cortisol response with moderate effects (ES_{sg} = 0.65). Secondary analyses focused on evaluating potential moderators that enhance the effectiveness of the V-TSST in eliciting a cortisol response. Sex, age, and immersivity all showed moderating influences on the effectiveness of V-TSST. Comparisons of the V-TSST with the traditional TSST and implications of the moderating variables are discussed in detail below.

The V-TSST has many advantages over the traditional TSST, including conservation of resources and increased standardization. With the current study results, it is also shown to elicit a cortisol response with a medium effect size ($ES_{sg} = 0.65$). This suggests that V-TSST may be a promising alternative to the traditional TSST for future stress research. However, a recent meta-analysis evaluated 186 studies to determine the effectiveness of the traditional TSST and indicated that it was effective in eliciting a cortisol response with a large effect size ($d = 0.925$; Allen et al., 2014). Thus, the V-TSST, although effective, may not be equal in effectiveness to the traditional TSST in regards to eliciting a cortisol response. However, based on the moderator analyses, several factors seem to have a differential impact on the effect size of V-TSST.

4.1. Moderator analyses: Implications and future research

The moderator analysis of sex demonstrated significant differences in effect sizes when analyzed both as a continuous and as a categorical variable. Male only studies showed a significantly higher effect size than mixed sex or female only studies. This is unsurprising, given the level of evidence in the traditional TSST literature which consistently indicates that men tend to have greater overall cortisol reactivity than women (Liu et al., 2017). These results may implicate that V-TSST induces cortisol reactivity in men at a large effect size similar to that of the traditional TSST, whereas for mixed groups or for women only, the V-TSST induces cortisol reactivity at a moderate effect size. Future V-TSST studies should incorporate greater mixed-sex and/or women only participants and follow the traditional TSST research by looking at effects of menstrual cycle and oral contraceptive use on female cortisol reactivity to the V-TSST, as these are known to significantly alter diurnal cortisol patterns and may impact the pattern of results obtained (Kirschbaum et al., 1999).

When examining age as a moderating variable, most studies were conducted with participants who were in their 20 s. Analysis of age as a continuous variable failed to show significant moderation effects, but there were significant differences in effect sizes when studies were divided at a mean age over or under 25 years old, as that is the generally agreed upon age at which brain maturation is complete (Arain et al., 2013). The under 25 group had significantly greater effect sizes than the over age 25 group. Based on this analysis, it could be hypothesized that individual below the threshold of brain maturation may find the V-TSST more stressful. Alternatively, younger populations may experience greater cortisol reactivity to V-TSST due to the ubiquity of and familiarity with technology in younger populations. They are more likely to have at least some familiarity with virtual reality as it has become more popular in recent years. However, due to the lack of studies looking at older populations and the fact that age did not have a moderating effect as a continuous variable, it is difficult to draw any firm conclusions on the effects of age, but it does generate interesting questions for future V-TSST research. Direct comparisons between youth, middle aged, and older populations could elucidate any differences in stress reactivity and the believability of each type of V-TSST.

Perhaps the most interesting moderator analysis was that which divided V-TSST studies into immersive or non-immersive virtual environments. In particular, the divide between immersive virtual environments that completely replaced both audio and visual inputs from the physical world with audio and visual inputs from the virtual environment were interesting compared to non-immersive environments. This group included studies that used CAVE environments and immersive HMDs. Studies that used immersive virtual environments had a significantly larger effect size than studies that used non-immersive virtual environments. This difference suggests that immersive virtual environments may be more effective in eliciting a cortisol response. This could be because the immersive virtual environments may induce a greater sense of ‘being there’ and feel more plausible to participants, qualities which may contribute to more realistic behavior in virtual environments (Slater, 2019). Although the sample size in the current meta-analysis precluded further investigation of what particular virtual environment characteristics may be predictive of effects to cortisol in response to acute stress, future meta-analytic research of V-TSST with a larger sample of studies should investigate this question as it may yield informative results.

4.2. Limitations and outlook

Several important limitations currently exist in the V-TSST research. Most notably, there are limitations with the sample sizes and population characteristics of extant studies. Many of the studies included in this review have fairly small sample sizes and almost half are comprised of only male participants. Given that there is evidence that men tend to have higher cortisol to the TSST than women (Liu et al., 2017), the bias towards male only participants in the reviewed studies may have considerable influence on overall effect size. Additional V-TSST studies with mixed-sex samples are needed to determine if the effectiveness of V-TSST changes with the addition of more female participants. Also, of importance for population characteristics, only four of the studies in this meta-analysis reported participant race/ethnicity percentages. This was not enough data to perform any moderator analyses based on race or ethnicity, and future researchers employing the TSST in any capacity should report this subject characteristic. Another limitation is the potential of V-TSST to be cost-prohibitive in regards to required equipment. The moderator analysis of immersivity indicates that more immersive environments show greater effect sizes than less immersive environments. However, most of the immersive environments in the reviewed studies consisted of a CAVE system, which is expensive and not feasible for many research laboratories. Two of the immersive VEs consisted of more cost-effective HMDs, which are becoming more sophisticated and less expensive every year. While there are not currently enough studies to compare CAVE environments to HMDs, future research should consider comparing the two methods. If there are no cortisol reactivity differences between them, the field could potentially move towards more cost-effective virtual environment options in the future. This would align with the V-TSST goal of reducing resource-intensiveness, and it could further revolutionize the field of stress research by making V-TSST portable and easier to scale up. Portable V-TSST options could encourage a higher level of standardization to a variety of environments (*e.g.*, school-based research). Adopting V-TSST versions that use HMDs would also make it easier to scale up the research by using multiple HMDs and incorporating group TSST (TSST-G, von Dawans et al., 2011) into virtual environments. This would enable researchers to collect even more data at a higher level of standardization. As consumers continue to adopt recreation virtual environment products, the price of highly immersive virtual environments will drop, and utilizing highly-immersive V-TSST protocols may become an increasingly attractive option for laboratory-based research.

Based on *a priori* coding decisions, this meta-analysis also intended to look at the moderating influence of the type of confederate used on the V-TSST panel. Authors coded for whether each study used avatars (*i.e.*, computer generated virtual persons) or non-avatars (*i.e.*, pre-recorded video of real persons). Unfortunately, only two studies used real people on the confederate panel (Hawn et al., 2015; Kelly et al., 2007), which is not enough studies to draw any confident conclusions about the effect of avatar or non-avatar confederate panels. It would be interesting to see future research compare these different types of confederates in V-TSST to see whether this variable could have any effect on the discrepancy between V-TSST cortisol reactivity effect size and traditional TSST cortisol reactivity effect size.

For this meta-analysis, cortisol was the only stress reactivity variable that was considered as it was the most used physiological stress reactivity variable in the V-TSST literature. Future research may include other physiological measurement variables as these may be sensitive in similar or different ways than the results obtained for cortisol. The studies in this meta-analysis did not consistently use any other variable aside from cortisol, however several did look at changes in heart rate, which yielded a similar pattern of results. The average effect size across all studies for increase in heart rate from baseline to peak measurement was $ES_{sg} = 1.21$ ($SE = 0.09$). For individual studies, effect size calculations for heart rate increase ranged from medium to large (range = 0.59–1.74), and analysis of heterogeneity was significant ($Q = 19.56$, $df = 8$, $p = 0.01$), meaning that the variance between studies was greater than what may occur due to sampling error. Due to significant heterogeneity, we analyzed the heart rate data with a random-effects model rather than fixed-effects. However, since the number of V-TSST studies that reported heart rate data was fairly small ($n = 9$), we did not run moderator analyses to determine more specific reasons for the heterogeneity. Future research may consider investigating heart rate as perhaps a more robust or less intrusive measurement of physiological stress in response to V-TSST.

Finally, a more impactful meta-analysis could have been performed if each of the V-TSST studies also included a traditional TSST control group to compare cortisol reactivity. For this meta-analysis the within-groups standardized mean gain effect size was used because this direct comparison was not available for most studies. In fact, only two studies looked at direct comparisons of traditional TSST and V-TSST. The current meta-analysis represents a step towards implementing V-TSST more widely in the stress research literature by showing that V-TSST can reliably induce stress with moderate effects, but each current and newly created version of V-TSST should be thoroughly tested directly against the traditional TSST. As the traditional TSST has served the field of stress research for several decades and is extremely reliable in inducing stress, we do not advise researchers to blindly adopt any version of the V-TSST based on the results of this meta-analysis, particularly if it has not yet been compared directly with the traditional TSST.

4.3. Future recommendations

Based on the results of this meta-analysis and the limitations present in the current V-TSST literature, we have several recommendations for future V-TSST studies.

1 *Increase sample sizes and include and report more diverse population characteristics*. Additional V-TSST research needs to be conducted with greater sample sizes and more diverse sample populations, including mixed-sex populations with a variety of age ranges. Researchers should also report race/ethnicity percentages when studies are published to ensure potential stress reactivity differences in different populations can be evaluated.

2 *Use V-TSST virtual environments that are immersive*. The immersivity moderator analysis showed significantly greater effect sizes in immersive environments than non-immersive environments, and it is suggested that future researchers elect to use more immersive environments as a result.

3 *Include a traditional TSST control group*. Researchers should proceed with caution when using V-TSST and ensure that they also incorporate a traditional TSST control group for comparison. Given the limitations of this meta-analysis, and given that the overall effect size of V-TSST is still smaller than the traditional TSST, it is still advisable to include comparisons of the two in future studies.

4 *Examine a variety of physiological stress reactivity variables*. While cortisol is often used as the main indicator of stress reactivity, additional cardiovascular or endocrinological variables may provide useful information for stress reactivity to V-TSST.

5 *Consider answering some of the research questions posed in this meta-analysis*. V-TSST research could benefit from studies looking at (a) the reasons for lower levels of cortisol reactivity in mixed-sex groups and women only groups (*e.g.*, use of oral contraceptives), (b) cortisol reactivity to V-TSST for different age groups, (c) comparisons between different types of immersive VEs (*i.e.*, CAVE systems *vs* HMDs), and (d) comparisons between different types of confederates (*i.e.*, avatar *vs* non-avatar).

In conclusion, the finding that V-TSST elicits a statistically significant cortisol response of medium effect size provides quantitative evidence that V-TSST is effectively stressful. However, when comparing this finding with the cortisol response to the traditional TSST, it does not seem to be as effective. Additionally, immersivity of the virtual environment, along with participant sex and age significantly moderated the effectiveness of the V-TSST. There is a need for future V-TSST research to further elucidate these moderating effects and other potential moderators that could increase the effectiveness of V-TSST to the level of the traditional TSST.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not for profit sectors.

Declaration of Competing Interest

None.

References

Allen, A.P., Kennedy, P.J., Cryan, J.F., Dinan, T.G., Clarke, G., 2014. Biological and psychological markers of stress in humans: focus on the Trier Social Stress Test. Neurosci. Biobehav. Rev. 38, 94–124. https://doi.org/10.1016/j.neubiorev.2013.11.005.

Allen, A.P., Kennedy, P.J., Dockray, S., Cryan, J.F., Dinan, T.G., Clarke, G., 2016. The trier social stress test: principles and practice. Neurobiol. Stress 6, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001.

Annerstedt, M., Jönsson, P., Wallergård, M., Johansson, G., Karlson, B., Grahn, P., et al., 2013. Inducing physiological stress recovery with sounds of nature in a virtual reality forest—results from a pilot study. Physiol. Behav. 118, 240–250. https://doi.org/10.1016/j.physbeh.2013.05.023.

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., et al., 2013. Maturation of the adolescent brain. Neuropsychiatr. Dis. Treat. 9, 449–461. https://doi.org/10.2147/NDT.S39776.

Birkett, M.A., 2011. The trier social stress test protocol for inducing psychological stress. J. Visualized Exp.: JoVE(56). https://doi.org/10.3791/3238.

Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, É., Forget, H., et al., 2017. Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. Br. J. Psychiatry 210 (4), 276–283. https://doi.org/10.1192/bjp.bp.116.184234.

Bowman, D.A., McMahan, R.P., 2007. Virtual Reality: How Much Immersion Is Enough? Computer 40 (7), 36–43. https://doi.org/10.1109/MC.2007.257.

Chrousos, G.P., 2009. Stress and disorders of the stress system. Nat. Rev. Endocrinol. 5 (7), 374–381. https://doi.org/10.1038/nrendo.2009.106.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, N.J.: L. Erlbaum Associates.

Dickerson, S.S., Gruenewald, T.L., Kemeny, M.E., 2004. When the social self is threatened: shame, physiology, and health. J. Pers. 72 (6), 1191–1216. https://doi.org/10.1111/j.1467-6494.2004.00295.x.

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

Diemer, J., Mühlberger, A., Pauli, P., Zwanzger, P., 2014. Virtual reality exposure in anxiety disorders: impact on psychophysiological reactivity. World J. Biol. Psychiatry 15 (6), 427–442. https://doi.org/10.3109/15622975.2014.892632.

Drevon, D., Fursa, S.R., Malcolm, A.L., 2017. Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. Behav. Modif. 41 (2), 323–339. https://doi.org/10.1177/0145445516673998.

Duval, S., Tweedie, R., 2000. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication Bias in meta-analysis. Biometrics 56 (2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x.

Emmelkamp, P.M.G., 2005. Technological innovations in clinical assessment and psychotherapy. Psychoth. Psychosomat.; Basel 74 (6), 336–343.

Engert, V., Efanov, S.I., Duchesne, A., Vogel, S., Corbo, V., Pruessner, J.C., 2013. Differentiating anticipatory from reactive cortisol responses to psychosocial stress. Psychoneuroendocrinology 38 (8), 1328–1337. https://doi.org/10.1016/j.psyneuen.2012.11.018.

Fallon, M.A., Careaga, J.S., Sbarra, D.A., O'connor, M., 2016. Utility of a virtual trier social stress test: initial findings and benchmarking comparisons. Psychosom. Med. 78 (7), 835–840. https://doi.org/10.1097/PSY.0000000000000338.

Fich, L.B., Jönsson, P., Kirkegaard, P.H., Wallergård, M., Garde, A.H., Hansen, Å., 2014. Can architectural design alter the physiological reaction to psychosocial stress? A virtual TSST experiment. Physiol. Behav. 135, 91–97. https://doi.org/10.1016/j.physbeh.2014.05.034.

Foley, P., Kirschbaum, C., 2010. Human hypothalamus–pituitary–adrenal axis responses to acute psychosocial stress in laboratory settings. Neurosci. Biobehav. Rev. 35 (1), 91–96. https://doi.org/10.1016/j.neubiorev.2010.01.010.

Goodman, W.K., Janson, J., Wolf, J.M., 2017. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. Psychoneuroendocrinology 80, 26–35. https://doi.org/10.1016/j.psyneuen.2017.02.030.

Hawn, S.E., Paul, L., Thomas, S., Miller, S., Amstadterc, A.B., 2015. Stress reactivity to an electronic version of the Trier Social Stress Test: a pilot study. Front. Psychol. 6 (2016-21410-001).

Higgins, J., Green, 2011. Cochrane Handbook for Systematic Reviews of Interventions. Retrieved from. www.handbook.cochrane.org.

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

Jönsson, P., Wallergård, M., Österberg, K., Hansen, Å., Johansson, G., Karlson, B., 2010. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: a pilot study. Psychoneuroendocrinology 35 (9), 1397–1403. https://doi.org/10.1016/j.psyneuen.2010.04.003.

Kelly, O., Matheson, K., Martinez, A., Merali, Z., Anisman, H., 2007. Psychosocial stress evoked by a virtual audience: relation to neuroendocrine activity. Cyberpsychology Behav. 10 (5), 655–662. https://doi.org/10.1089/cpb.2007.9973.

Kirschbaum, C., Kudielka, B.M., Gaab, J., Schommer, N.C., Hellhammer, D.H., 1999. Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. Psychosom. Med. 61 (2), 154–162.

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.

Krijn, M., Emmelkamp, P.M.G., Olafsson, R.P., Biemond, R., 2004. Virtual reality exposure therapy of anxiety disorders: a review. Clin. Psychol. Rev. 24 (3), 259–281. https://doi.org/10.1016/j.cpr.2004.04.001.

Linninge, C., Jönsson, P., Bolinsson, H., Önning, G., Eriksson, J., Johansson, G., Ahrné, S., 2018. Effects of acute stress provocation on cortisol levels, zonulin and inflammatory markers in low- and high-stressed men. Biol. Psychol. 138, 48–55. https://doi.org/10.1016/j.biopsycho.2018.08.013.

Lipsey, M.W., Wilson, D.B., 2001. Practical Meta-analysis. Thousand Oaks, CA, US: Sage Publications, Inc.

Liu, J.J.W., Ein, N., Peck, K., Huang, V., Pruessner, J.C., Vickers, K., 2017. Sex differences in salivary cortisol reactivity to the Trier Social Stress Test (TSST): a meta-analysis. Psychoneuroendocrinology 82, 26–37. https://doi.org/10.1016/j.psyneuen.2017.04.007.

McEwen, B.S., 2007. Physiology and neurobiology of stress and adaptation: central role of the brain. Physiol. Rev. 87 (3), 873–904. https://doi.org/10.1152/physrev.00041.2006.

Montero-López, E., Santos-Ruiz, A., García-Ríos, M.C., Rodríguez-Blázquez, M., Rogers, H.L., Peralta-Ramírez, M.I., 2018. The relationship between the menstrual cycle and cortisol secretion: daily and stress-invoked cortisol patterns. Int. J. Psychophysiol. 131, 67–72. https://doi.org/10.1016/j.ijpsycho.2018.03.021.

Orwin, R.G., 1983. A fail-safe N for effect size in meta-analysis. J. Educ. Stat. 8 (2), 157–159. https://doi.org/10.2307/1164923.

Powers, M.B., Emmelkamp, P.M.G., 2008. Virtual reality exposure therapy for anxiety disorders: a meta-analysis. J. Anxiety Disord. 22 (3), 561–569. https://doi.org/10.1016/j.janxdis.2007.04.006.

Rohatgi, A., 2018. (2018, January). WebPlotDigitizer (Version 4.1). Retrieved October 8, 2018, from. https://automeris.io/WebPlotDigitizer.

Rosenthal, R., 1979. The file drawer problem and tolerance for null results. Psychol. Bull. 638–641.

Santos-Ruiz, A., Garcia-Rios, M.C., Fernandez-Sanchez, J.C., Perez-Garcia, M., Muñoz-García, M.A., Peralta-Ramirez, M.I., 2012. Can decision-making skills affect responses to psychological stress in healthy women? Psychoneuroendocrinology 37 (12), 1912–1921. https://doi.org/10.1016/j.psyneuen.2012.04.002.

Santos-Ruiz, A.S., Peralta-Ramirez, M.I., Garcia-Rios, M.C., Muñoz, M.A., Navarrete-Navarrete, N., Blazquez-Ortiz, A., 2010. Adaptation of the Trier Social Stress Test to virtual reality: Psycho-physiological and neuroendocrine modulation. J. Cyberther. Rehab. 3 (4), 405–415 (2011-02058-007).

Schneiderman, N., Ironson, G., Siegel, S.D., 2005. STRESS AND HEALTH: psychological, behavioral, and biological determinants. Annu. Rev. Clin. Psychol. 1, 607–628. https://doi.org/10.1146/annurev.clinpsy.1.102803.144141.

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. Int. J. Psychophysiol. 110, 47–55. https://doi.org/10.1016/j.ijpsycho.2016.10.008.

Slater, M., 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos. Trans. Biol. Sci. 364 (1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138.

Stanković, S., 2015. Virtual reality and virtual environments in 10 lectures. Synth. Lect. Image Video Multimed. Process. 8 (3), 1–197. https://doi.org/10.2200/S00671ED1V01Y201509IVM019.

Viechtbauer, W., 2017. metafor: Meta-Analysis Package for R (Version 2.0-0). Retrieved from. https://CRAN.R-project.org/package=metafor.

von Dawans, B., Kirschbaum, C., Heinrichs, M., 2011. The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. Psychoneuroendocrinology 36 (4), 514–522. https://doi.org/10.1016/j.psyneuen.2010.08.004.

Zänkert, S., Bellingrath, S., Wüst, S., Kudielka, B.M., 2018. HPA axis responses to psychological challenge linking stress and disease: What do we know on sources of intra- and interindividual variability? Psychoneuroendocrinology. https://doi.org/10.1016/j.psyneuen.2018.10.027.

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