Evaluating the Effectiveness of Graduated Stress Exposure in Virtual Spaceflight Hazard Training

Tor T. Finseth, **Nir Keren**, **Michael C. Dorneich**, **Warren D. Franke**, **Clayton C. Anderson**, and **Mack C. Shelley**, Iowa State University

Psychological and physiological stress experienced by astronauts can pose risks to mission success. In clinical settings, gradually increasing stressors help patients develop resilience. It is unclear whether graduated stress exposure can affect responses to acute stressors during spaceflight. This study evaluated psychophysiological responses to potentially catastrophic spaceflight operation, with and without graduated stress exposure, using a virtual reality environment. Twenty healthy participants were tasked with locating a fire on a virtual International Space Station (VR-ISS). After orientation, the treatment group (n = 10) practiced searching for a fire while exposed to a low-level stressor (light smoke), while the control group (n = 10) practiced without smoke. In the testing session, both groups responded to a fire while the VR-ISS unexpectedly filled with heavy smoke. Heart rate variability and blood pressure were measured continuously. Subjective workload was evaluated with the NASA Task Load Index, stress with the Short Stress State Questionnaire, and stress exposure with time-to-complete. During the heavy smoke condition, the control group showed parasympathetic withdrawal, indicating a mild stress response. The treatment group retained parasympathetic control. Thus, graduated stress exposure may enhance allostasis and relaxation behavior when confronted with a subsequent stressful condition.

Keywords: psychophysiology, resilience, space, stress exposure training, stress inoculation, virtual environments

Address correspondence to Michael Dorneich, Iowa State University, 3028 Black Engineering, 2529 Union Dr., Iowa State University, Ames, IA 50011, USA, dorneich@iastate.edu.

Journal of Cognitive Engineering and Decision Making
2018, Volume 12, Number 4, December 2018, pp. 248–268
DOI: 10.1177/1555343418775561
Article reuse guidelines: sagepub.com/journals-permissions

Several uncontrollable and unpredictable emergencies have occurred aboard the International Space Station (ISS) including fire, depressurization, and toxic environments (Summers, Johnston, Marshburn, & Williams, 2005). Astronauts are responsible for either remedying the situation or being prepared to evacuate the station. These situations can be highly stressful, since the consequences of not responding appropriately to the situation can be catastrophic. To prepare for emergency situations, NASA astronauts train approximately 40 hr using a full-scale ISS mock-up at NASA Johnson Space Center. Emergency training is a small, but crucial, piece of the overall training process, necessitating rigorous scheduling and international travel over the course of 2 years. Given the sheer volume of operational spaceflight training that astronauts undergo, it is challenging to incorporate training interventions focused on remaining calm during potentially life-threatening situations.

Stress arises in transactional situations where the individual's perceived demands tax or exceed the perceived coping resources, which can result in negative physiological, psychological, behavioral, or social outcomes (Lazarus & Folkman, 1984). The individual's appraisal of a situation determines the extent to which the situation is stressful. When stress occurs, the process of allostasis is the body's attempt to adapt, maintain, or regain stable levels of functioning (McEwen, 2001; McEwen & Wingfield, 2003). With appropriate training, a healthy allostatic state can be maintained during exposure to intense acute stress. However, if allostasis is not maintained, outcomes can include a disruption in information processing (Gaillard, 2001), impaired or rigid decision making (Ellis, 2006; Starcke & Brand, 2012), and declines in cognitive performance (Lieberman et al., 2005), potentially leading to freezing behavior or tonic immobility (Abrams, Carleton, Taylor, & Asmundson, 2009) and an impairment in performance during the crisis (Delahaij, Gaillard, & Soeters, 2006).

Most spaceflight training is performed intensively and frequently, but it is primarily focused on mastering tasks. Over time, repetitively practiced skills become automated, thereby requiring less attention and being more resistant to disruption (Driskell, Salas, Johnston, & Wollert, 2008). However, stress can negatively affect performance even with high levels of task training (Orasanu & Backer, 1996). Overlearned skills may not lead to effective coping, except in situations where the stressors are well-known (Delahaij et al., 2006). When individuals are exposed to unpredictable situations, new stressors, or radical environmental changes, a maladaptive response may occur. Learning how to respond to unpredictable acute stress could be helpful in spaceflight. While task training focuses on the automaticity of the task itself, stress training focuses on reduction of the stress response through coping. Developing coping strategies decreases the potential for negative cognitive, psychological, and behavioral reactions (Leipold & Greve, 2009; Meichenbaum & Cameron, 1989; Serino et al., 2014). Moreover, during acute stress, coping mechanisms increase the likelihood of staying calm and relatively relaxed.

Stress inoculation training (SIT) could potentially help astronauts stay calm by building resilience to acute, sequential, and chronic stressors (Meichenbaum, 1985). The SIT approach is a three-phased flexible form of cognitive behavioral therapy. The initial phase of training is conceptual education focusing on the nature of stress and stress effects. The second phase involves acquisition of coping skills and consolidation of skills already possessed, where a variety of coping skills are rehearsed in preparation for stressful situations. The final phase of SIT can include application of these coping skills across multiple inoculation sessions with increasingly demanding levels of stressors (Meichenbaum, 2017; Saunders, Driskell, Johnston, & Salas, 1996).

Resilience can be achieved when the individual's appraisal promotes protective coping without experiencing mental health disruptions, despite being subject to stressors (Fletcher & Sarkar, 2013; Kalisch, Müller, & Tüscher, 2015). With the goal of improving resilience, SIT is commonly used for the prevention and management of stress. Stress management differs from stress prevention in that the former focuses on reactive care and support after a stressful incident, while the latter focuses on proactive measures to reduce the stress response (Staal, 2004). SIT has been validated as a stress management intervention for individuals in chronically stressful work settings (Foley, Bedell, LaRocca, Scheinberg, & Reznikoff, 1987; Perna, Antoni, Baum, Gordon, & Schneiderman, 2003) or with anxiety, depression, or posttraumatic stress disorder (Hourani et al., 2011). While SIT is often used in clinical therapy, more research is needed to explore its utility with healthy individuals experiencing acute stress (Rose et al., 2013).

One component of clinical SIT, graduated stress exposure, provides a mechanism for becoming comfortable with stress by providing trainees with a heightened sense of control and competency (Keinan & Friedland, 1996). Graduated stress exposure could be readily integrated into astronaut training. Here, stressors could be introduced during task training with the stress levels gradually increasing during a training session, or over a series of sessions, to promote control of the individual's threat appraisal (Fornette et al., 2012; Johnston & Cannon-Bowers, 1996). A single session may be sufficient for stress response improvement (Baumann, Gohm, & Bonner, 2011), although multiple sessions may be better at fostering confidence in preparation for realistic stress levels.

In clinical therapy, the flexible framework of SIT allows the training to be modified to fit a patient's needs. However, additional research is needed on stress training for healthy people working in challenging environments, such as astronauts (Rose et al., 2013). Limited evidence exists to guide trainers in the selection of effective SIT intervention techniques (Crawford, Wallerstedt, & Khorsan, 2013; Regehr, Glancy, & Pitts, 2013); thus, research is needed to determine how the separate components of SIT contribute differentially to appraisal, biological arousal, and training effectiveness during occupational tasks (Robson & Manacapilli, 2014; Saunders et al., 1996).

Therefore, the purpose of this study is to assess the extent to which the individual contribution of graduated exposure to a task-specific stressor affects the physiological response to a critical spaceflight hazard. Specifically, we hypothesize that exposure to a light level of virtual smoke, during training focused on responding to a fire threat on the ISS, would improve the psychophysiological responses to a subsequent simulated emergency with heavy-smoke exposure, in comparison to a group that was trained without smoke exposure during the prior training. Collectively, this study serves as a proof-of-concept that this aspect of SIT may be worthy of consideration for inclusion in future astronaut training. The novel contribution of this paper is to provide insight into the effectiveness of graduated stress exposure as a component of SIT on modifying the response to stress and whether the brief exposure to an acute spaceflight hazard, in healthy people, improves psychological and physiological responses to subsequent exposure.

METHODS

General Experimental Design

Study participants came to the laboratory three times. The first visit was an orientation tour of the ISS in a virtual reality environment (VR-ISS). Here, participants learned the emergency response procedure to a fire, practiced navigating the VR-ISS, and completed a written quiz indicating their way finding and emergency response abilities. The second visit, the training session, focused on completing a fire response protocol in the VR-ISS; half the participants were exposed to light smoke during this session, while the other half were not. The third visit, the testing session, was another fire response protocol; however, both groups were exposed to an unexpected and rapid accumulation of heavy smoke during the session. Participants were randomly assigned using a 1:1 ratio to either the treatment (exposed to low level of smoke in prior session) or control (no exposure to smoke in prior session) group for the two fire drill sessions. The location of the source of the smoke was the same for both participant groups (i.e., treatment and control) but varied between the training and testing sessions. The purpose of changing the source location was to prevent a learning effect between the training and testing session. The level of task difficulty in terms of required procedures and source location was kept constant across sessions and groups. Physiological responses and psychological states were assessed in both sessions. All study procedures were approved by the Iowa State University Institutional Review Board.

Participants

Potential participants were excluded if they reported having severe anxiety, claustrophobia, pregnancy, simulation sickness, seizures, heart abnormalities, circulatory problems, or implanted electromagnetic devices. Twenty-two subjects consented, but two withdrew prior to finishing the experiment. The final sample was 20 adult males, with a mean age of 22.5 years (SD = 2.2), from the Iowa State University community. None of the participants had prior experience with VR. The demographics were as follows: 60% Caucasian, 15% African American, 15% Hispanic or Latino, and 10% Asian or Asian American.

Task/Scenarios

All participants were asked to follow a simplified ISS emergency fire response procedure in the VR-ISS with the goal of locating the source of the smoke. The simulation followed the NASA ISS emergency fire procedures which contained instructions for crew responsibilities, air contaminants location sampling, and ISS system configuration (NASA, 2013). During the simulation, smoke was generated from a source in one of the modules aboard the ISS U.S. Orbital Segment (see Figure 1). Participants began the simulation in the Node 1 module, since this is the "safe haven," closest to the Russian operations segment and the Soyuz escape capsule on the ISS (see U.S. Orbital Segment interior of the ISS in Figure 1).

The simulation took place in the C6, a virtual reality room at Iowa State University; Figure 2a show participants in the VR-ISS in the C6. Figure 2b is an example of a view that participants saw in the simulation, including the location of a fireport.

To aid detection and location of the source of smoke, participants experienced different virtual smoke and corresponding atmospheric contaminant levels based upon the treatment or control

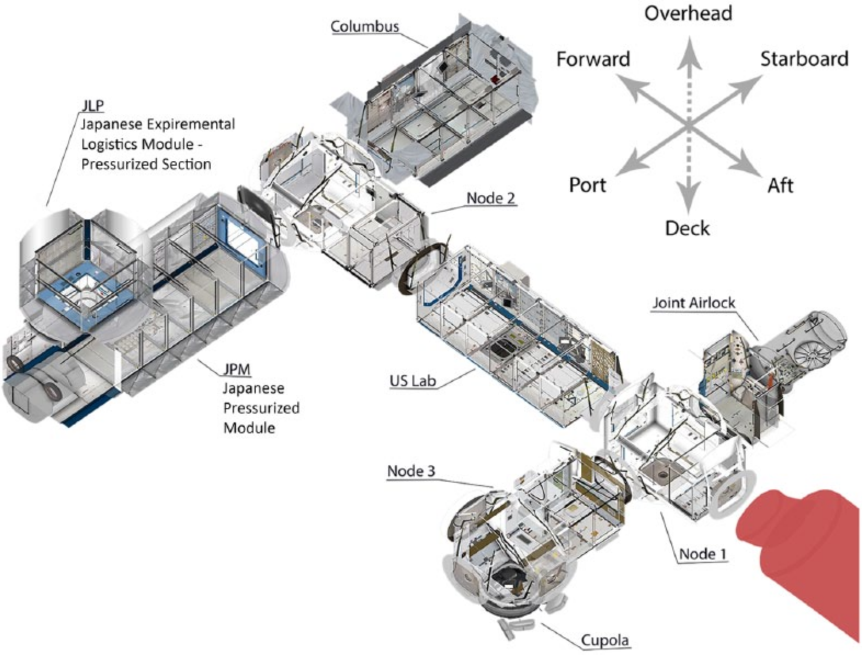

Figure 1. Simulated International Space Station (ISS) configuration. The Russian segment of the ISS was not included in the simulation.

Figure 2. The (a) virtual reality International Space Station (VR-ISS) and (b) a fireport.

group to which they were assigned (i.e., training with light smoke or no smoke). Atmospheric contaminant levels rose as a function of time and distance from the fire source. However, the smoke density changed as a function of time; therefore, participants could not rely solely on visual smoke cues to detect the source. Participants evaluated contaminant levels using a hand-held joystick programmed to emulate the NASA-used Compound Specific Analyzer–Combustion Products (CSA-CP) device. The purpose of the CSA-CP on board the ISS is to determine the level of atmospheric contaminants that are expected to be released due to potential fire, specifically dictating the length of time before the Protective Breathing Apparatus is required. Participants were instructed to assess the atmospheric state by using the CSA-CP to display the levels of carbon monoxide (CO), hydrogen chloride (HCl), and hydrogen cyanide (HCN) in parts per million (Figure 3). Upon participant command, a floating window appeared in front of the participant with the contaminants' concentration values. The window disappeared after 5 seconds. Based upon the participant's recall of the previously assessed contaminant

Figure 3. The Compound Specific Analyzer–Combustion Products (CSA-CP) in the virtual reality environment (VRE) displays environmental noxious gas readings (left) and fireport with identifier code (right).

levels in each VR-ISS module, the highest reading would indicate the approximate location of the source of smoke.

Once the participants identified the ISS module where the fire source was located, they began sampling fireports within the module to locate the "rack" that caused the fire. The VR-ISS includes fireport labels accurately placed on the racks throughout the ISS (Figure 2). The labels have a unique code identifier, which includes the module name, module surface, and rack number. Participants were trained on the fireport identifier codes during the first laboratory visit. The simulation ended when participants identified and reported the fireport label on the individual module rack, which had the highest CSA-CP reading, or when the simulation smoke became condensed to a level where visibility was almost zero (which occurred 10 min from the beginning of the simulation), at which point the experiment controller stopped the experiment.

Procedure

The experiment was divided into three laboratory visits, each lasting approximately 60 min (Figure 4). The second and third laboratory visits, consisting of the experimental sessions, occurred at least 24 hr apart (M = 39 hr, SD = 35). The first laboratory visit served to (a) educate and orient the participant to the ISS through virtual reality environment (VRE) practice and (b) develop task skills necessary to perform a simple fire response procedure. Participants were trained on the ISS layout and modules, how to navigate using module labels (e.g., PORT = left side, STBD = right side), and how to identify key landmarks within the modules (e.g., locations of the treadmill and cupola). The participants were then trained on the ISS fire procedures, which included equipment, fireport rack labeling (e.g., "JPM1F3"), and proper procedure responses. Participants were given a guided acclimation walkthrough in the VR-ISS, which included reiteration of the ISS layout, navigation, landmarks, and operation of the VRE. At the end of this visit, a written test was used to confirm participants' ability to navigate the VR-ISS and perform the emergency fire procedure.

During the second visit, termed the training session, participants completed the emergency procedure in either a light smoke or no smoke condition (Figure 4) based on their assignment to the treatment or control group, respectively. Participants were not informed of their group assignment, only that they would be performing the emergency fire response in the VR-ISS. Prior to entering the VR-ISS, the participants were given a brief review of the ISS layout and navigation. Before the session began, participants completed the Short State Stress Questionnaire (SSSQ) to assess subjective stress levels. Participants then sat quietly for 10 min while baseline physiological data were collected. They then entered the VR-ISS and completed the fire response procedure. Postsession SSSQs were completed after finishing the simulation.

In the third visit, termed the testing session, participants completed the same emergency response procedure as in training. However, both the treatment and control groups were exposed to higher ("heavy smoke") levels of virtual smoke and atmospheric contaminants. Participants were unaware that the smoke levels

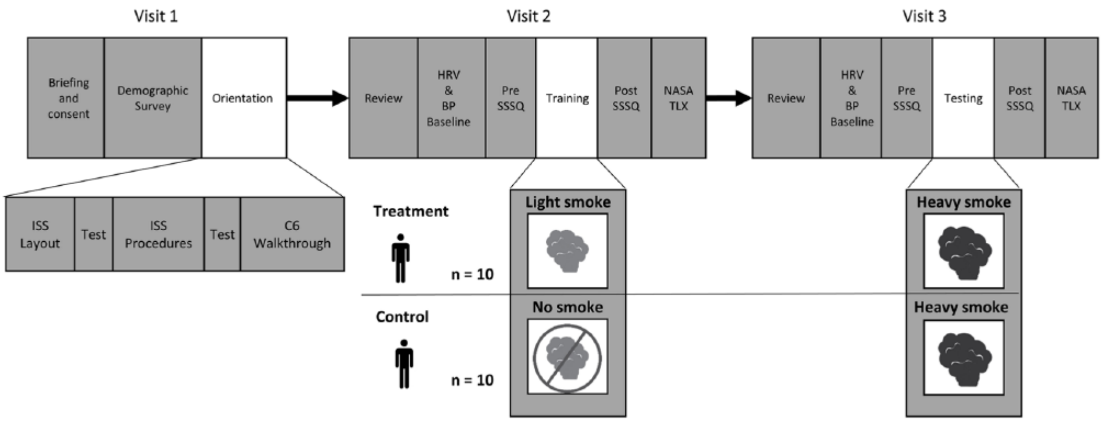

Figure 4. Design and procedure of the study. BP = blood pressure; HRV = heart rate variability; ISS = International Space Station; SSSQ = Short State Stress Questionnaire; TLX = Task Load Index.

would increase significantly. The review, questionnaires, and physiological measurements were administered as described for the training visit. At the end of the experiment, participants were debriefed on the implementation of the heavy smoke conditions.

Independent Variables

There are two independent variables in the experiment: session (training, testing) and group (control, treatment). Participants were placed into one of two groups: (1) a treatment group with prior light smoke exposure (training) followed by heavy smoke (testing), and (2) a control group with no smoke exposure (training) followed by heavy smoke (testing).

Dependent Variable Measures

The study used both psychological and physiological indices of stress: perceived subjective stress state and psychophysiological biomarkers of the stress response. It has been recommended that training in VREs should measure stress through multiple standardized avenues such as cognitive perceived state, mediating factors, or psychophysiological biomarkers (Serino et al., 2014). Thus, data collected in the experiment included autonomic nervous system (ANS) responses and cognitive workload. The dependent variables are summarized in Table 1.

ANS responses. Two overlapping branches of the ANS, the sympathetic nervous system branch and parasympathetic nervous system branch, determine the arousal or restorative functions targeting organs in response to a stressor. The overall "stress response" of the heart and vascular system is a result of the interplay between these two branches. It includes the effects of locally secreted neurotransmitters (e.g., norepinephrine) as well as systemic modulators (e.g., epinephrine). In essence, the sympathetic nervous system primes the body for action, while the parasympathetic nervous system regulates organ and gland functions during rest.

The ANS responses to stress were assessed with two measures: heart rate variability (HRV) and blood pressure (BP). HRV is comprised of time domain and frequency indices that reflect the balance between ANS-mediated relaxation and arousal (Hourani et al., 2011; Malik, 1996). Heart rate data were collected via electrocardiogram (ECG, modified CS_5 lead configuration). The ECG was sampled at 2,048 Hz using Biopac MP150 hardware and recorded using AcqKnowledge software (Version 3.8.2, Biopac Systems Inc.). Spectral analysis of ECG was performed using the Matlab-based toolbox Kubios HRV software (Niskanen, Tarvainen, Ranta-Aho, & Karjalainen, 2004). The raw data were first inspected visually for artifacts and corrected using the Kubios artifact correction option through detection of successive ECG wave peaks (RR intervals); the default low artifact correction level of Kubios was used for detecting RR intervals differing "abnormally" from

TABLE 1: Description of Dependent Variable Metrics, Units, and Frequencies

Dependent Variable	Metric	Components	Association	Unit	Measurement Frequency
Autonomic stress response	Heart rate variability (HRV)	HR	Cardiac activity	BPM	Baseline before session, throughout each session
		HF n.u.	Parasympathetic (i.e., vagal activity)	Dimensionless	
		LF n.u.	Sympathetic and parasympathetic	Dimensionless	
		LF/HF ratio	Sympathovagal balance	Dimensionless	
		RMSSD	Parasympathetic	*Ms*	
		pNN50	Parasympathetic	%	
Autonomic stress response	Blood pressure	Systolic (SBP) and		mmHg	Before session, throughout sessions
		diastolic (DBP) blood pressure		mmHg	
Psychological stress response	Short Stress State Questionnaire (SSSQ)	Engagement, distress, worry		Likert scale	Pre- and postsession
Workload	NASA Task Load Index (TLX)			Likert scale	Postsession
Time-to-completion				Min	Throughout each session

Note. BPM = beats per minute; HF = high frequency; LF = low frequency; pNN50 = proportion of the number of pairs of successive normal R-waves that differ by more than 50 msec; RMSSD = root-mean-square difference of successive normal R-wave intervals.

the local mean RR interval (Tarvainen & Niskanen, 2012). First-order trend correction was applied. Spectral density analysis of the HRV was used to parse the data into a low-frequency (LF) (0.04–0.15 Hz) band reflecting sympathetic activity with vagal modulation and a high-frequency (HF) (0.15–0.4 Hz) band reflecting parasympathetic activity. The very-low-frequency (VLF, <0.04 Hz) band was not included in this study because it is unreliable for short-term recordings (<5 min) (Malik, 1996). The LF and HF components were normalized to their total power in order to remove influences of VLF and the influence of changes in total power that may occur with autonomic arousal (e.g., HF/[HF + LF] × 100). The LF/HF ratio was calculated to assess sympathovagal balance, which is an index of the relative amount of sympathetic activity (the extent at which the individual is hyperaroused for action; sometimes referred to as “fight or flight response”) relative to parasympathetic activity (the extent to which an individual feels at ease) of the ANS. Therefore, this ratio concisely represents the individual’s physiological stress response. Time domain analysis of the ECG was performed to quantify the amount of variance in the interbeat interval through the root-mean-square difference of successive normal R-wave intervals (RMSSD) and the proportion of the number of pairs of successive normal R-waves that differ by more than 50 msec (pNN50). Both RMSSD and pNN50 represent vagal control within the time domain and are correlated to HF power (Shaffer, McCraty, & Zerr, 2014). The HRV time domain and frequency bands for each participant were calculated in 60-sec intervals over the duration of each session. The first minute of the data was omitted to prevent anticipatory stress responses from skewing the assessments.

Systolic BP (SBP) and diastolic BP (DBP) were collected as another measure of cardiovascular reactivity. DBP and SBP can reflect changes in the total peripheral resistance of blood vessels. Increases in local sympathetic activity cause constriction of blood vessels, while reductions in sympathetic activity (or more parasympathetic activity) lead to dilation. In the absence of changes in cardiac output, decreases in blood vessel constriction are usually reflected by decreases in DBP. In the present study, beat-to-beat BP data were collected. A finger cuff was placed on the participants' nondominant hand over the middle phalanx of either the long or ring finger (Finapres 2300; Ohmeda). The nondominant arm was placed in an arm sling to standardize the position of the hand relative to the heart between all participants. Data were recorded at 1,024 Hz. After instrumentation and before each session, participants sat quietly for 10 min while baseline physiological data were collected. To calibrate the finger cuff, an oscillometric noninvasive BP cuff (CNAP Monitor 500, CNSystems Medizintechnik AG) was placed on the participant's dominant upper arm, and BP was measured twice during this 10-min baseline period. The two systems showed similar BP measurements, suggesting the arm sling sufficiently accounted for potential hydrostatic pressure differences between the fingers and heart level. The raw data were inspected visually for artifacts and corrected using AcqKnowledge software. BP values were saved in 15-sec interval samples. To give ample time for a resting state to occur and to prevent anticipatory stress interference, baseline BP data were calculated as the mean of the data from Minutes 5 to 8 of the 10-min baseline. The mean baseline value was subtracted from the session data to determine change scores (Zhang & Han, 2009). The first minute of the data was omitted to prevent any anticipatory stress interference.

Psychological stress responses. The SSSQ was administered before and after each laboratory visit to assess multiple dimensions of the subjective response to stressful environments. The SSSQ assesses three state factors: task engagement, distress, and worry (Helton, 2004). Engagement refers to qualities of energetic arousal, motivation, and concentration. Distress is characterized by feelings of tense arousal, hedonic tone, and confidence control. Worry relates to self-focus, self-esteem, and cognitive interference (Matthews et al., 1999). The three-factor SSSQ scale scores for pre- and postsession were calculated for each participant. The factor scores from both pre- and postsession were standardized against normative means and standard deviation values from a large sample of British participants obtained by Matthews et al. (2002) and using the method of Helton and colleagues (Helton, 2004; Helton, Matthews, & Warm, 2009). Average difference scores for each state measure were then calculated for treatment and control groups. These were used to calculate the absolute difference between sessions, resulting in a z score.

Workload. The NASA Taskload Index (TLX) was used to assess the subjective workload of the task during exposure (Hart & Staveland, 1988). The NASA TLX measures six dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration level. The NASA TLX was administered after the completion of a session. Participant scores on the six numerical rating scales were computed in the 0 to 100 range and as an unweighted participant mean for each of the six dimensional subscales (Nygren, 1991).

Time-to-complete. The time-to-complete the fire response procedure was used to assess the stressor duration between participants. Since the stressor intensity (i.e., smoke density) increased over time, the length of time spent completing the procedure could influence the stress response.

Materials

The research was conducted in the C6, the high resolution virtual reality room at the Virtual Reality Applications Center (VRAC) at Iowa State University. The room is a 10 ft. × 10 ft. × 10 ft. cube in which all six screens have projected 4K stereoscopic images that provide total immersion in a virtual world. VirtuTrace, a full-featured "experiment engine," allows researchers to develop immersive experiment protocols for display in a fully immersive system (see Keren, Franke, Bayouth, Harvey, & Godby, 2013). Users moved in the virtual environment by taking a step in the desired direction, whereupon the system would track standing body position to facilitate motion through the environment. The VR-ISS was created using NASA-provided models of the U.S. Orbital Segment interior of the ISS (Figure 1). The Russian segment of the ISS was not included in the simulation since a model of this segment was not available.

Data Analysis Plan

Data analysis was performed using SPSS software (Version 23.0; IBM Corp.). For comparison of HRV components and BP, a linear mixed model (LMM) was used to calculate the fixed effect interaction of Group × Session. Random effect from participant sampling was used in the covariance matrix. All HRV metrics and HR were Winsorized to 3 *SD*s to reduce the impact of outliers. The LF/HF variables had a moderate positive skew and was Log(x+1) transformed to normalize the data for parametric analysis. Subjective stress questionnaires (SSSQ) and workload questionnaires (TLX) were checked for normality and subsequently assessed using the nonparametric Mann–Whitney U test. Results were considered significantly different at the $p \leq .05$ level. A statistical trend is defined as results with $.05 < p < .10$. All results shown are means (M) and standard error (SE).

Effect sizes were calculated for the fixed effects and interaction effects. Cohen's (1988) d was used for the standardized effect size in units of standard deviation. Cohen's d effect size guidelines are reported as small for $.2 < |d| < .5$, medium for $.5 < |d| < .8$, and large for $|d| > .8$ (Cohen, 1988). Effect sizes for the Mann–Whitney U tests were calculated with normal approximation z to r. Cohen's guidelines for Pearson correlation r score effect size are adopted as small for $.1 < r < .3$, medium for $.3 < r < .5$, and large for $r > .5$ (Fritz, Morris, & Richler, 2012).

To assess whether the standard deviations of the interaction differed between the group and session, a 95% confidence interval (CI) was generated by the parametric bootstrap resampling technique (Efron & Tibshirani, 1993). Parametric bootstrapping uses a fitted model based on the experiment sample data to generate synthetic data through replication, which yields a sampling distribution for a larger population. The bootstrap population distribution can then provide a robust empirical CI estimation. Some studies have recommended at least 2,000 replications for CI (DiCiccio & Efron, 1996; Efron, 1987); the present study used 10,000 replications for greater CI accuracy. Bootstrapping was performed using R software (Version 3.3.2; R Foundation for Statistical Computing). The mixed model in R was verified by comparison of maximum likelihood estimation in SAS software (Version 9.2; SAS Institute).

RESULTS

ANS Stress Response by HRV and Heart Rate

Comparison of baseline physiological measures between groups revealed no significant differences. Physiological data during the VR-ISS procedures are presented in Tables 2 and 3. The main effects of group and session were not significant for HR, but differences were seen with the HRV variables.

A main effect was seen for session, where LF/HF was significantly higher for testing (M = 0.6, SE = 0.042) compared to training (M = 0.50, SE = 0.041), $F(1, 58) = 8.13$, $p = .006$, $d = -0.70$. A significant increase was found for normalized LF for testing (M = 69.8, SE = 2.8) compared to training (M = 63.0, SE = 2.7), $F(1, 60) = 9.41$, $p = .003$, $d = -0.79$. In contrast, a significant decrease was seen for normalized HF for testing (M = 30.0, SE = 2.8) compared to training (M = 36.8, SE = 2.7), $F(1, 59) = 9.47$, $p = .003$, $d = 0.79$. The HRV time domain indices also changed. RMSSD was significantly decreased for testing (M = 41.9, SE = 3.1) compared to training (M = 48.7, SE = 3.0), $F(1, 46) = 7.14$, $p = .01$, $d = 0.72$, and pNN50 was significantly decreased for testing (M = 14.3, SE = 2.4) compared to training (M = 18.4, SE = 2.4), $F(1, 35) = 8.56$, $p = .006$, $d = 0.54$. No significant main effects for group were found for any HRV metrics, but as described below, several Multiple Group × Session interaction effects were seen.

The interaction effect for normalized HF of HRV during testing showed a statistical trend

TABLE 2: Descriptive Statistics for the LMM Measures of HRV and HR, *M* (*SE*)

	Control		Treatment	
Metric	Training	Testing	Training	Testing
Log(LF/HF + 1)	0.51 (0.059)	0.66 (0.062)	0.51 (0.056)	0.54 (0.057)
LF n.u.	63.3 (3.9)	74.3 (4.12)	62.6 (3.7)	65.3 (3.76)
HF n.u.	36.5 (3.88)	25.5 (4.1)	37.1 (3.68)	34.5 (3.75)
HR (BPM)	83.2 (4.05)	86.4 (4.08)	84.4 (4.01)	82.1 (4.02)
RMSSD (msec)	47.9 (4.34)	35.8 (4.13)	49.4 (4.13)	48.0 (4.20)
pNN50 (%)	18.6 (3.40)	10.7 (3.58)	18.2 (3.33)	17.9 (3.36)

Note. BPM = beats per minute; HF = high frequency; HR = heart rate; HRV = heart rate variability; LF = low frequency; LMM = linear mixed model; pNN50 = proportion of the number of pairs of successive normal R-waves that differ by more than 50 msec; RMSSD = root-mean-square difference of successive normal R-wave intervals.

TABLE 3: Significance Level and Effect Size for the LMM Measures of HRV and HR

Metric	$P_{\text{Fixed effect}}$ (Group)	$P_{\text{Fixed effect}}$ (Session)	$P_{\text{Interaction}}$ (Group × Session)	Cohens *d* (Interaction Effect Only)
Log(LF/HF + 1)	0.455	0.006*	0.051	–0.49
LF n.u.	0.343	0.003*	0.066	–0.48
HF n.u.	0.345	0.003*	0.062	0.49
HR (BPM)	0.786	0.765	0.092	–0.30
RMSSD (msec)	0.232	0.01*	0.042*	0.57
pNN50 (%)	0.465	0.006*	0.011*	0.50

Note. BPM = beats per minute; HF = high frequency; HR = heart rate; HRV = heart rate variability; LF = low frequency; LMM = linear mixed model; pNN50 = proportion of the number of pairs of successive normal R-waves that differ by more than 50 msec; RMSSD = root-mean-square difference of successive normal R-wave intervals.
*$p \leq .05$.

(i.e., $.05 < p < .10$) for increasing in the treatment group compared to the control group, $F(1, 59) = 3.61$, $p = .062$, with a small-moderate effect size ($d = 0.49$) (Figure 5). The normalized LF had a decreasing trend for the treatment group compared to the control group, $F(1, 60) = 3.51$, $p = .066$, $d = -0.48$ (Figure 6). Likewise, the interaction effect for the LF/HF responses due to differences in session approached statistical significance, $F(1, 58) = 3.97$, $p = .051$, $d = -0.49$ (Figure 7). The interaction effect for HR also showed a trend, $F(1, 19) = 3.15$, $p = .092$, $d = -0.30$ (Figure 8). The interaction effect for the time domain indices of RMSSD was significantly higher for the treatment group compared to the control group, $F(1, 46) = 4.39$, $p = .042$, $d = 0.56$ (Figure 9), and the pNN50 was significantly higher for the treatment group compared to the control group, $F(1, 35) = 7.25$, $p = .011$, $d = 0.50$ (Figure 10).

The Group × Session interaction effect CIs for a large population distribution, calculated via parametric bootstrap, showed that the original sample estimate was highly reliable when the original sample model was assessed for 10,000 replications. The interaction effects for LF/HF, RMSSD, and pNN50 were significant. The interaction effects for normalized LF, normalized HF, and HR showed a statistical trend (Table 4).

ANS stress response by BP. From the data presented in Tables 5 and 6, no significant group or session effect was found. The DBP response for the Group × Session interaction effect illustrated in Figure 11 was not significant, $F(1, 40)$

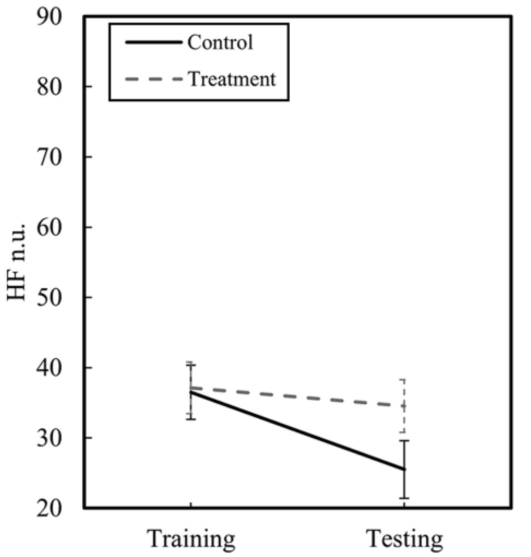

Figure 5. Mean and standard error of HF (n.u.). HF = high frequency.

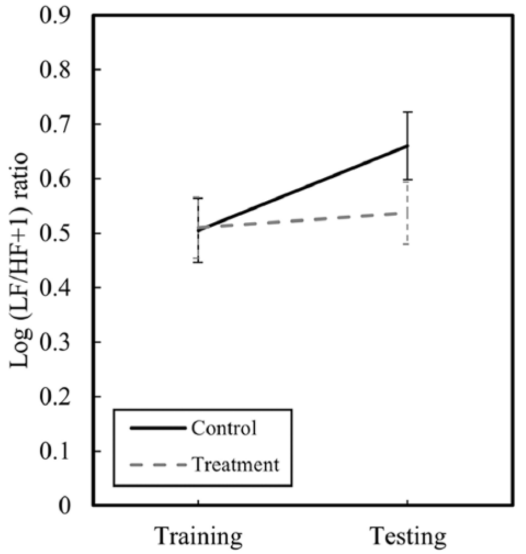

Figure 7. Mean and standard error of Log(LF/HF + 1) ratio. HF = high frequency; LF = low frequency.

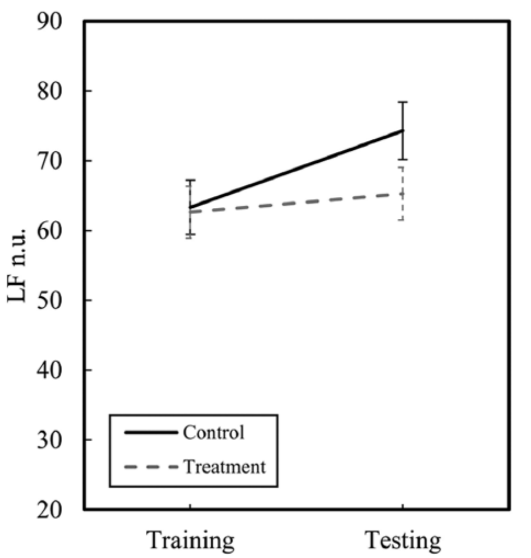

Figure 6. Mean and standard error of LF (n.u.). LF = low frequency.

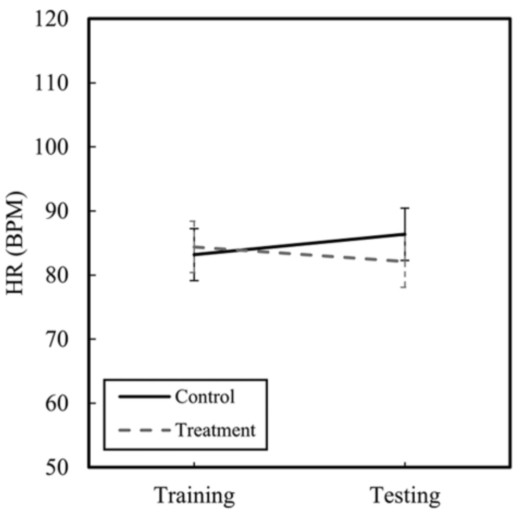

Figure 8. Mean and standard error of HR. BPM = beats per minute; HR = heart rate.

= 1.91, p = .174, d = .32. Similarly, no significant difference was found with the SBP interaction effect, F(1, 26) = .043, p = .836, d = –.06 (Figure 12). Parametric bootstrap of the Group × Session interaction effect shown in Table 7 confirmed that DBP and SBP would remain unchanged even in a larger population sample.

Stress State

The main effect of group and the main effect of session were not significant for the SSSQ scale factors. However, the simple main effect for the task engagement factor during the training session was significantly less for the treatment (light smoke) condition (M = 0.22, SE = 0.13) than for the control (no smoke) condition

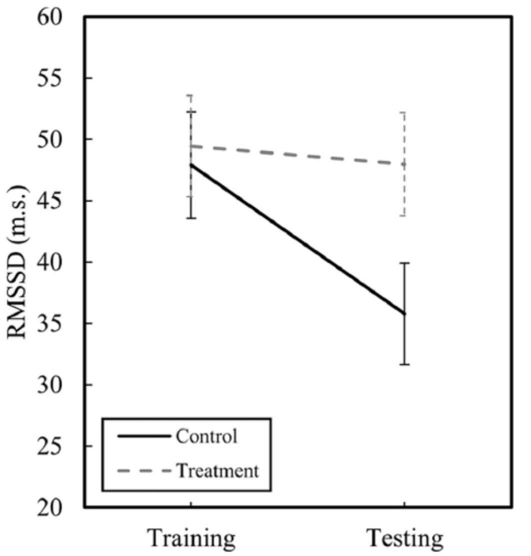

Figure 9. Mean and standard error of RMSSD. RMSSD = root-mean-square difference of successive normal R-wave intervals.

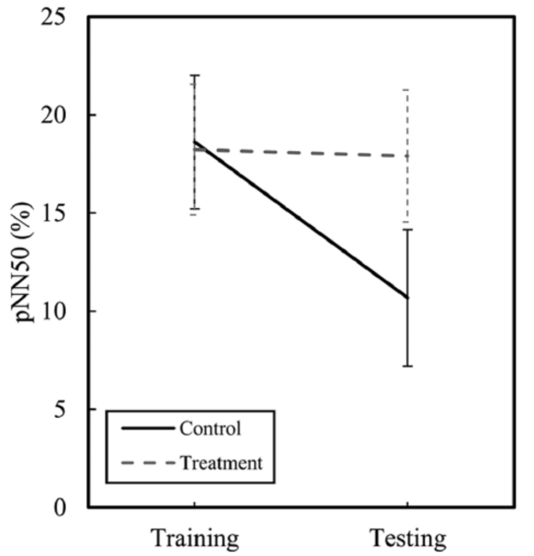

Figure 10. Mean and standard error of pNN50. pNN50 = proportion of the number of pairs of successive normal R-waves that differ by more than 50 msec.

(M = 0.46, SE = 0.071) (U = 23.5, p = .041, r = –.46). The simple main effect in the testing session was not significantly different between the treatment (M = 0.41, SE = 0.14) and control (M = 0.34, SE = 0.11) (U = 47.5, p = .85, r = –.043).

Figure 13 illustrates the change in z score between the training and testing for both the control and treatment sessions for task engagement, distress, and worry. The task engagement change score was not significantly different between the treatment group (M = 0.19, SE = 0.2) compared to the control group (M = –0.12, SE = 0.12) (U = 31, p = .137, r = –.33). The distress change score was not significantly different between the treatment group (M = 0.095, SE = 0.093) compared to the control group (M = 0.095, SE = 0.067) (U = 48, p = .876, r = –.035). The worry change score was not significantly different between the treatment group (M = 0.32, SE = 0.23) compared to the control group (M = 0.055, SE = 0.097) (U = 43.5, p = .621, r = –.11). All other sources of variance in the analysis lacked statistical significance or trends (p > .10).

Workload (NASA-TLX)

Figure 14 illustrates the workload scores for both training and testing. No significant differences were found for the group main effect. The main effect of session on temporal demand indicated that testing (Mdn = 76.2) had significantly greater temporal demand than training (Mdn = 49.3) (U = 114.5, p = .02, r = –.37). The main effect of session on physical demand indicated that testing (Mdn = 28.6) has greater physical demand than training (Mdn = 23.8) (U = 127.5, p = .048, r = –.31). When evaluating the TLX within a single session, simple main effects for group were not significant.

Time-to-Complete

The main effect for group on the time-to-complete was significantly higher for the treatment (M = 7.29, SE = 0.66) than for the control group (M = 5.32, SE = 0.66), $F(1, 18)$ = 4.43, p = .05, with a large effect size (d = 1.19). A main effect trend was detected for session, such that the time to complete the emergency procedure was higher for training (M = 6.73, SE = 0.52) than for testing (M = 5.89, SE = 0.52), $F(1, 18)$ = 3.30, p = .086, d = 0.51. No significant interaction was found.

DISCUSSION

The purpose of this study was to assess the extent to which graduated exposure to a task-specific stressor affected the physiological

TABLE 4: Estimate of Interaction Effect with Bootstrap 95% Confidence Intervals

	Estimate	Lower Bound	Upper Bound	Bootstrapped *p* Value (Group × Session)
Log(LF/HF + 1)	0.27	0.011	0.55	0.047*
LF n.u.	–8.21	–16.5	1.08	0.067
HF n.u.	8.28	–0.76	16.8	0.065
HR (BPM)	–6.46	–13.8	1.14	0.091
RMSSD (msec)	12.8	1.57	23.6	0.023*
pNN50 (%)	7.80	2.14	13.6	0.008*

Note. BPM = beats per minute; HF = high frequency; HR = heart rate; LF = low frequency; pNN50 = proportion of the number of pairs of successive normal R-waves that differ by more than 50 msec; RMSSD = root-mean-square difference of successive normal R-wave intervals.
*$p \leq .05$.

TABLE 5: Descriptive Statistics for the LMM Measures of BP, *M* (*SE*)

	Control		Treatment	
	Training	Testing	Training	Testing
SBP (mmHg)	25.2 (1.04)	21.8 (1.28)	26.4 (0.73)	27.0 (1.60)
DBP (mmHg)	11.0 (0.49)	9.05 (0.80)	11.1 (0.47)	14.1 (0.64)

Note. BP = blood pressure; DBP = diastolic blood pressure; LMM = linear mixed model; SBP = systolic blood pressure.

TABLE 6: Inferential Statistics for the LMM Measures of BP, *p* Values and Effect Size of Interaction

	$P_{\text{Fixed effect}}$ (Group)	$P_{\text{Fixed effect}}$ (Session)	$P_{\text{Interaction}}$ (Group*Session)	Cohens *d* (Interaction Effect)
SBP (mmHg)	0.44	0.87	0.84	–0.06
DBP (mmHg)	0.31	0.40	0.17	0.32

Note. BP = blood pressure; DBP = diastolic blood pressure; LMM = linear mixed model; SBP = systolic blood pressure.

response to a critical spaceflight hazard. Specifically, we hypothesized that exposure to a light level of virtual smoke, during training that is focused on responding to a fire threat on the ISS, would attenuate the psychophysiological responses to a subsequent simulated emergency with an unexpected heavy smoke condition. Compared to the control group, the autonomic responses of the treatment group suggested that they had relatively less activation of the sympathetic nervous system, enhanced allostasis, and adaptability to a stress response. These results suggest that prior exposure to a low-level stressor attenuates the sympathovagal response to a more stressful condition of the same task in virtual reality. Moreover, these psychophysiological measures suggest that using only one component of the SIT framework, graduated exposure to a stressor, can positively affect the responses to exposure to a more severe version of the stressor (i.e., testing).

When evaluating the ability to stay calm as indicated by the ANS stress response, the unchanged state of the treatment group's normalized HF component, RMSSD, and pNN50 suggests that participants retained parasympathetic

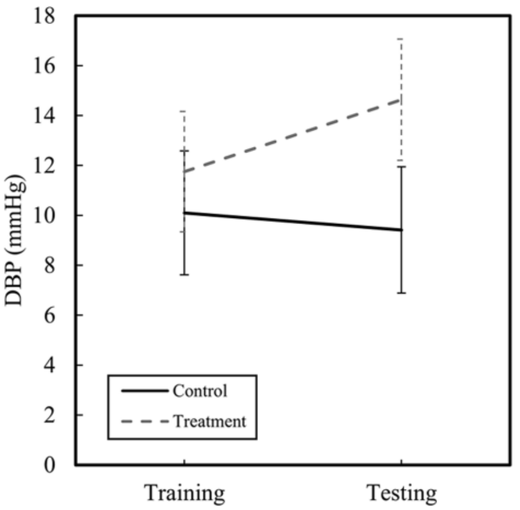

Figure 11. Mean and standard error of diastolic blood pressure (DBP) from baseline.

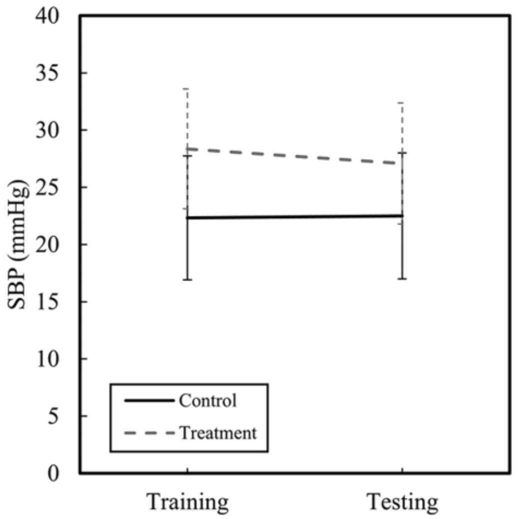

Figure 12. Mean and standard error of systolic blood pressure (SBP) from baseline.

modulation. In contrast, the control group's parasympathetic withdrawal is characteristic of a mild stress response. This response suggests they were more stressed and relatively unable to maintain a calm behavioral state (Shaffer et al., 2014). The control group was relatively unaware of this change, as evident by a lack of between-group differences in their SSSQ scores.

Any changes seen in the HRV were primarily meditated by withdrawal of the parasympathetic branch of the ANS (Porges, 1995) leading to relatively greater sympathetic control as indicated by the decrease in normalized HF and increase in normalized LF (i.e., comparing Figures 5 and 6). This interpretation is consistent with Hjortskov et al. (2004), who found that short lasting exposure to psychosocial stressors indicated parasympathetic withdrawal and that an unchanged sympathetic activity was responsible for the increase in LF/HF ratio. Further, associations between resting autonomic balance and psychological resilience are supported by the polyvagal theory (Porges, 1995) in which the primary response to a stressor is mediated by the parasympathetic nervous system component of the ANS (Lewis et al., 2015). Collectively, these findings suggest that without prior exposure to a mild threat, participants were unable to relax when exposed to the more severe threat, thus contributing to greater autonomic arousal and higher stress levels.

The results reinforce the use of HRV, rather than HR, as a tool for measuring psychophysiological stress responses. Results showed a weakly elevated HR for the control and decrease for the treatment group between the sessions. However, the trend is small and only a subtle reflection of the ANS activity (Shaffer et al., 2014). Similarly, BP remained unchanged, suggesting no change in vasoconstriction during the stress response. While HR and BP can be perceivable indications of stress at times, the nonreaction during a stressful situation verifies the usefulness of HRV to detect stress without relying on human subjectiveness.

The control group's SSSQ task engagement during training was higher than the treatment group. Due to the control group not experiencing smoke exposure in training, the absence of physiological stress may have resulted in a sense of control and challenge-related appraisal, and therefore higher levels of engagement for the control group. Challenge appraisal and effective coping have been shown to be associated with higher parasympathetic activation (Geisler, Kubiak, Siewert, & Weber, 2013; Laborde, Lautenbach, & Allen, 2015). The treatment group's unchanged engagement between sessions possibly reflects the parasympathetic activation, indicating that inoculation fostered a challenge perspective and coping. However, the control group's parasympathetic withdrawal

TABLE 7: Estimate of Interaction Effect With Bootstrap 95% Confidence Intervals

	Estimate	Lower Bound	Upper Bound	Bootstrapped *p* Value (Group × Session)
SBP (mmHg)	–0.54	–18.3	18.2	0.955
DBP (mmHg)	3.84	–3.93	11.7	0.334

Note. DBP = diastolic blood pressure; SBP = systolic blood pressure.

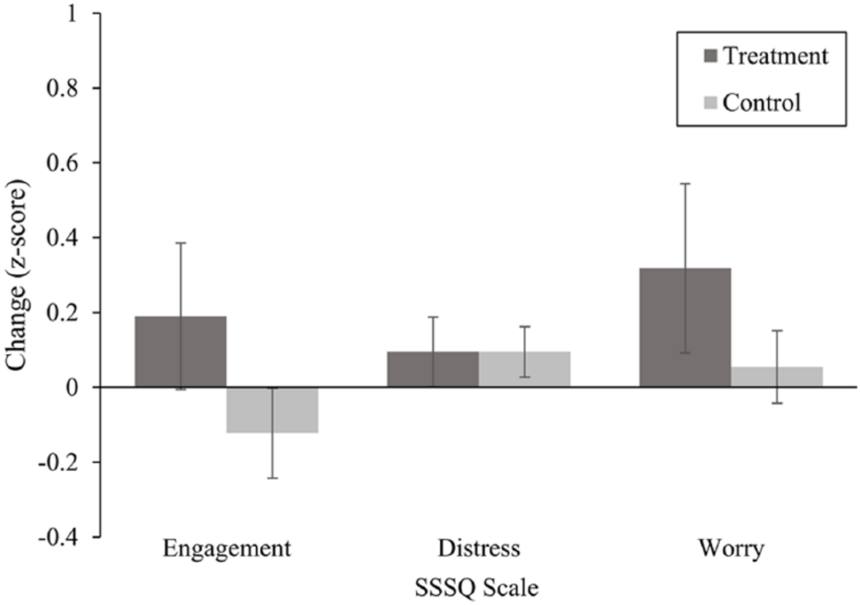

Figure 13. Mean and standard error of Short State Stress Questionnaire (SSSQ) change scores between sessions.

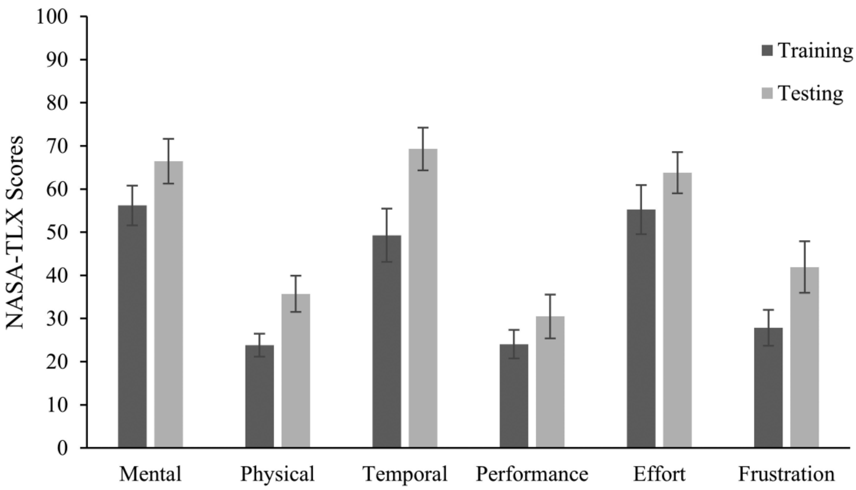

Figure 14. Workload profile for session main effects (mean and standard error). TLX = Task Load Index.

during the testing session suggests the challenge became a threat appraisal and subsequently resulted in stress (Schwerdtfeger & Derakshan, 2010). While the absence of stress during training may have been relatively beneficial in the short term for the control group participants, it did little to prepare them for the more severe testing.

The subjective assessment of task workload showed no change between groups but changed between sessions for the temporal and physical workload dimensions. These results are reinforced by the high SSSQ engagement results for both groups. Previous studies have suggested that task engagement has moderate correlations to the temporal and physical dimensions of the NASA TLX but also has high correlations to TLX's mental demand and effort (Matthews, Szalma, Panganiban, Neubauer, & Warm, 2013). Possible explanations for the increased temporal and physical response to testing are the increase in perceived time pressure of locating a fire and the greater physical fatigue associated with the stress response evoked by the heavy smoke scenario. The unchanged state of mental demand and effort between sessions can be explained by the already relatively high mental demands of practicing the emergency fire procedure during the training session. The procedure requires navigation and spatial knowledge as well as the short-term memory task of remembering containment levels to support any movement decisions. Navigational processes include visual attention, spatial memory, and working memory operations, which feed information to executive function decision making (Bajaj et al., 2008).

In stress training interventions that involve performing tasks, it may be difficult to distinguish whether the stressor or the task is primarily influencing HRV. One could question whether changes in HRV between sessions may be partly attributed to the change in workload rather than the exposure of the experimental stressors. Changes in task complexity and workload influence HRV (Jorna, 1992), which is more pronounced in executive-level functions such as sustained attention (Luque-Casado, Perales, Cárdenas, & Sanabria, 2016). Similarly, HRV is correlated to the stress caused by evaluations of threat and safety and inhibition of unwanted memories and intrusive thoughts (Shaffer et al., 2014; Thayer, Ahs, Fredrikson, Sollers, & Wager, 2012). Based on this interconnectedness, both the perceived threat caused by the stressors and the executive control required by the task are theorized to be feedback and feedforward in nature, allowing both functions to simultaneously influence regulation of ANS (Thayer, Hansen, Saus-Rose, & Johnsen, 2009). Therefore, without a clear understanding of the extent to which HRV measurements are influenced by task or stressor, stress response improvements (lower HRV) might be due to task learning effect from improved executive control rather than resilience or an increased ability to relax during the stressor. To de-confound this problem, we can compare changes in task engagement and changes in workload across sessions and groups.

Executive control has been shown to be associated with SSSQ task engagement (Matthews & Zeidner, 2012). Executive functions require short-term memory, focused attention, or manipulation of new information (Thayer et al., 2009). In this task, the executive functions are remembering the NASA emergency procedures while maintaining sustained attention to fire-related cues and navigating the VR-ISS to detect and locate the fire. The progression from the training to testing session showed an increase in temporal and physical workload, while the SSSQ task engagement remained unchanged. However, the ANS activities reflected by HRV showed a large parasympathetic decrease for the control group during the experiment, but no change for the treatment group (see Figures 9 and 10). The workload demands increased from the training session to the testing session even though there was no change in the task engagement, which suggests that the executive function caused by the task was not responsible for the stress response. Collectively, the lack of SSSQ task engagement differences and the presence of HRV differences between groups in the testing session suggest that the perceived threat from the simulated environment (i.e., scenario change from light/no smoke to heavy smoke) was the primary source of stress for participants, without direct influences of stress elicited from either the changing workload or the procedures.

In the present study, high workload levels and a decreased engagement between trials for the control group would have provided evidence that experiencing a more stressful situation without inoculation may result in an overload of attentional resources. As complexity increases for memory and search tasks, allocation of free resources can be impaired and decrease the task engagement (Matthews & Davies, 2001). The impairment is due to visual search tasks and spatial information using common mechanisms in working memory (Woodman & Luck, 2004). We would have expected that the simulated threat would have interfered with executive function more in the control group than in the treatment group because of the larger difference in the manipulated smoke levels (from training to testing). Therefore, we would have expected to see the task engagement decrease from training to testing in the control group but remain unchanged from training to testing for the treatment group. However, the task engagement from the training to testing session for the control group did not rise to the level of significance ($p = .137$; see Figure 13).

The time to complete the procedure was 1 min longer for treatment compared to control but also 2 min longer for training compared to testing. Although the treatment group took 2 min longer to complete the task compared to the control group, the LF/HF power spectra during the training session suggests no difference in the stress response. Both groups completed the task quicker in the testing session than in the training session, most likely due to familiarization with the procedure. While the time to complete the procedure differed between group and session, the participants were only told they would be practicing the emergency procedure and not informed that a rapid emergency response would be evaluated or was more favorable. In other words, "performance" was not emphasized as a desired attribute. The longer completion time for the treatment group during training might have resulted from a decrease in visibility during the task due to the increase in virtual smoke density.

It is important to address the reduced emphasis on performance. One could consider time-to-completion as a performance metric that we would expect would decrease with increased decision making. Keren et al. (2013) studied firefighters' decision making under stress in a VRE. The results indicated that time-to-decision was significantly longer with veteran firefighters when compared to novice firefighters, potentially indicating novices out performed veterans. However, the investigation demonstrated that veterans used their experience to better assess the time window available for a response, which in turn they used to enhance their situational awareness and to better size up the situation at hand, rather than debating the decision task. Thus, an increase in speed did not reflect better performance. Postexperiment decision analysis revealed that veterans' decision quality was higher. Bayouth (2011) analyzed firefighters' decision performance under two different stressors: (a) difficult tradeoff and (b) time pressure. When under time pressure, time-to-decision did not vary between veterans and novices. The quality of veterans' decision quality was less prevalent when under time stress rather than under difficult tradeoffs. Thus, performance is a difficult, complex attribute to assess.

Several factors may limit the generalizability of our conclusions. The participant sample size was adequate to measure psychophysiological measures, but a larger sample size would likely increase the reliability of our subjective stress measurements. The participant sample included males only; thereby potential gender effects could not be detected. The goal of this study was to assess the effect of stress training for healthy people working in challenging environments; however, future research on the efficacy of using stress training for specific tasks would benefit from a sample more closely related to the relevant occupational demographic, such as astronauts. While stress appraisal and coping have substantial variability between individuals, task proficiency can lead to a heightened sense of control and mitigate stress (Orasanu & Backer, 1996). Therefore, because an astronaut sample may be more proficient at emergency procedures than a college student sample, the effect of stress training may not sufficiently generalize to the representative population. An experiment with a sample with similar education, age, and gender that reflects the astronaut population would strengthen the generalizability of the findings.

Although the graduated stress exposure elicited an improved physiological response after only two sessions, it remains uncertain whether more sessions will affect the benefits. In a review by Saunders et al. (1996), the beneficial effects of SIT increase as the number of training sessions increase. A single session can be sufficient for tempering the stress response, but five to seven sessions produce a robustly positive effect. Further, the present study only evaluated the application phase of SIT (third phase) and intentionally omitted the two phases focused on stress education and acquisition of coping skills. Participants were not given any instruction in these two areas, and their coping abilities were not assessed prior to the experiment. Individual differences in coping abilities could potentially have affected our results. Several limitations also exist for the HRV analysis and interpretation of results. There is an influence of respiration rate on the cardiac cycle, which limits clear interpretation of the HF spectrum as an index of cardiac vagal tone. RMSSD and pNN50 may be correlated to HF power, however the influence of respiration rate on the indices is uncertain (Shaffer et al., 2014). Further, there are concerns that the LF/HF ratio cannot be considered as an index of sympathovagal balance due to assumptions of vagal and sympathetic nerve traffic to the heart (Eckberg, 2000).

CONCLUSION

This work addresses a gap in the literature with respect to enhancing the resilience of healthy individuals to acute stress, which may be more pronounced in the realm of spaceflight. The current findings suggest that even modest graduated stress exposure shows promise as a useful tool during spaceflight procedure training to prepare for emergencies. Spaceflight is dangerous, and much of the current training paradigm for astronauts is focused on task exposure and mastery. However, the present study suggests that psychophysiological responses to life-threatening situations can be mitigated by graduated stress exposure training and this can be done concurrent with task-specific training. The results suggest that participants who received prior exposure to a stressor enhanced their ability to remain calm during an emergency procedure task in a virtual ISS. Moreover, this study shows that graduated exposure training is a beneficial training component in SIT.

While not the focus of this study, the results suggest that a VRE system for training astronauts to handle stressful situations would likely increase their ability to cope and perform under acute highly stressful situations. The use of a VRE to administer gradual stress levels proved to be effective at eliciting an appropriate stress response to varying conditions. VREs offer a safe and controlled environment for training for traumatic or hazardous situations. Simultaneously, VREs have potential to solve two common stress training problems—namely, treatment consistency as well as reconciling differences between the training environment and the environment in which the task is performed (Meichenbaum, 2017). To date, VRE simulations for intravehicular activity have been used far less during NASA training in lieu of full-scale mock-ups (Aoki, Oman, Buckland, & Natapoff, 2008; Gancet, Chintamani, & Letier, 2012). Since a high-fidelity VRE utilizes less resources than traditional fire training using mock-ups, emergency training in VREs may be a suitable alternative for NASA astronauts in preparation for some aspects of missions to the ISS.

Future work is needed to study further the inoculation effects of using a stress training pedagogy for spaceflight applications. The current study did not explore whether the effects of SIT enhance performance for spaceflight operations, as has been theorized for other occupations (e.g., McClernon, McCauley, O'Connor, & Warm, 2011). In addition, although preliminary findings indicate that graduate exposure can attenuate relaxation mechanisms, future research should investigate the use of VR for all phases of SIT used in a preventative approach. Robson and Manacapilli (2014) note that when SIT is implemented in the full three-phases, current use of VR during SIT primarily occurs in the application phase, and this use is in the context of the full three-phased implementation of SIT. To administer SIT in its entirety in VR, a phased training approach with the three phases separated minimizes the interference of stressors affecting new trainees trying to learn skills (Friedland & Keinan, 1992). Although our intent was to evaluate graduated stress exposure in isolation, the full SIT framework in VR may have more pronounced improvements in stress response.

ACKNOWLEDGMENTS

The authors are grateful to the NASA Johnson Space Center ER7's Virtual Reality Laboratory and Integrated Graphics Operations and Analysis Laboratory for providing the 3D models of the ISS. The authors thank Kevin Godby for his assistance in integrating the ISS models into VirtuTrace and for running VirtuTrace during the sessions and also thank Dr. Elizabeth Shirtcliff for offering expertise on the human stress response.

ORCID iD

Michael Dorneich https://orcid.org/0000-0001-6386-4787

REFERENCES

Abrams, M. P., Carleton, R. N., Taylor, S., & Asmundson, G. J. (2009). Human tonic immobility: Measurement and correlates. *Depression and Anxiety*, *26*(6), 550–556. doi:10.1002/da.20462

Aoki, H., Oman, C., Buckland, D., & Natapoff, A. (2008). Desktop-VR system for preflight 3D navigation training. *Acta Astronautica, 63*(7–10), 841–847. doi:10.1016/j.actaastro.2007.11.001

Bajaj, J., Hafeezullah, M., Hoffmann, R., Varma, R., Franco, J., Binion, D., Hammeke, T., & Seian, K. (2008). Navigation skill impairment: Another dimension of the driving difficulties in minimal hepatic encephalopathy. *Hepatology*, *47*(2), 596–604. doi:10.1002/hep.22032

Baumann, M. R., Gohm, C. L., & Bonner, B. L. (2011). Phased training for high-reliability occupations: Live-fire exercises for civilian firefighters. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 548–557. doi:10.1177/0018720811418224

Bayouth, S. T. (2011). *Examining firefighter decision making process and choice in virtual reality* (Doctoral dissertation). Ames, IA: Iowa State University Digital Repository.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillside, NJ: Lawrence Earlbaum Associates.

Crawford, C., Wallerstedt, D., & Khorsan, R. (2013). A systematic review of biopsychosocial training programs for the self-management of emotional stress: Potential applications for the military. *Evidence-Based Complementary and Alternative Medicine*, *2013*, 747694. doi:10.1155/2013/747694

Delahaij, R., Gaillard, A.W. K., & Soeters, J. M. L. M. (2006). Stress training and the new military environment. In *Human dimensions in military operations—Military leaders' strategies for addressing stress and psychological support* (Meeting Proceedings RTO-MP-HFM-134, Paper 17A, pp. 17A-1–17A-10). Neuilly-sur-Seine, France: NATO Reasearch and Technology Organisation.

DiCiccio, T., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*(3), 189–212.

Driskell, J. E., Salas, E., Johnston, J. H., & Wollert, T. N. (2008). Stress exposure training: An event-based approach. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (pp. 271–286). London: Ashgate.

Eckberg, D. L. (2000). Physiological basis for human autonomic rhythms. *Annals of Medicine*, *32*(5), 341–349.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, *82*(397), 171–185.

Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, *57*, 1–436.

Ellis, A. (2006). System breakdown: The role of mental models and transactive memory in the relationship between acute stress and team performance. *Academy of Management Journal*, *49*(3), 576–589. doi:10.5465/AMJ.2006.21794674

Fletcher, D., & Sarkar, M. (2013). Psychological resilience: A review and critique of definitions, concepts, and theory. *European Psychologist*, *18*(1), 12–23. doi:10.1027/1016-9040/a000124

Foley, F. W., Bedell, J. R., LaRocca, N. G., Scheinberg, L. C., & Reznikoff, M. (1987). Efficacy of stress-inoculation training in coping with multiple sclerosis. *Journal of Consulting and Clinical Psychology*, *55*(6), 919–922. doi:10.1037/0022-006X.55.6.919

Fornette, M.- P., Bardel, M.- H., Lefrançois, C., Fradin, J., Massioui, F., & Amalberti, R. (2012). Cognitive-adaptation training for improving performance and stress management of air force pilots. *The International Journal of Aviation Psychology*, *22*(3), 203–223. doi:10.1080/10508414.2012.689208

Friedland, N., & Keinan, G. (1992). Training effective performance in stressful situations: Three approaches and implications for combat training. *Military Psychology*, *4*(3), 157–174. doi:10.1207/s15327876mp0403_3

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology, General*, *141*(1), 2–18. doi:10.1037/a0024338

Gaillard, A. W. K. (2001). Stress, workload, and fatigue as three biobehavioral states: A general overview. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 623–640). Mahwah, NJ: Lawrence Erlbaum.

Gancet, J., Chintamani, K., & Letier, P. (2012, September). *Force feedback and immersive technologies suit (FITS): An advanced concept for facility-less astronaut training*. Paper presented at the International Symposium on Artificial Intelligence, Robotics and Automation in Space, Turin, Italy.

Geisler, F., Kubiak, T., Siewert, K., & Weber, H. (2013). Cardiac vagal tone is associated with social engagement and self-regulation. *Biological Psychology*, *93*(2), 279–286. doi:10.1016/j.biopsycho.2013.02.013

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183. doi:10.1016/S0166-4115(08)62386-9

Helton, W. (2004). Validation of a Short Stress State Questionnaire. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(11), 1238–1242.

Helton, W. S., Matthews, G., & Warm, J. S. (2009). Stress state mediation between environmental variables and performance: The case of noise and vigilance. *Acta Psychologica*, *130*(3), 204–213. doi:10.1016/j.actpsy.2008.12.006

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, *92*(1–2), 84–89. doi:10.1007/s00421-004-1055-z

Hourani, L. L., Kizakevich, P. N., Hubal, R., Spira, J., Strange, L. B., Holiday, D. B., Bryant, S., & McLean, A. N.(2011). Predeployment stress inoculation training for primary prevention of combat-related stress disorders. *Journal of CyberTherapy & Rehabilitation*, *4*(1), 101–116.

Johnston, J., & Cannon-Bowers, J. (1996). Training for stress exposure. In E. Salas (Ed.), *Stress and human performance* (pp. 223–256). Mahwah, NJ: Lawrence Erlbaum.

Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology, 34*(2–3), 237–257. doi:10.1016/ 0301-0511(92)90017-O

Kalisch, R., Müller, M. B., & Tüscher, O. (2015). A conceptual framework for the neurobiological study of resilience. *The Behavioral and Brain Sciences*, *38*, e92. doi:10.1017/ S0140525X1400082X

Keinan, G., & Friedland, N. (1996). Training effective performance under stress: Queries, dilemmas, and possible solutions. In J. E. Driskell & E. Salas (Eds.), *Stress and human performance* (pp. 257–277). Mahwah, NJ: Lawrence Erlbaum.

Keren, N., Franke, D. W., Bayouth, S. T., Harvey, E. M., & Godby, K. M. (2013). VirtuTrace: Training for making decisions under stress in virtual environments. In *Proceedings of the Annual Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, FL.

Laborde, S., Lautenbach, F., & Allen, M. S. (2015). The contribution of coping-related variables and heart rate variability to visual search performance under pressure. *Physiology & Behavior*, *139*, 532–540. doi:10.1016/j.physbeh.2014.12.003

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York, NY: Springer-Verlag.

Leipold, B., & Greve, W. (2009). Resilience: A conceptual bridge between coping and development. *European Psychologist*, *14*(1), 40–50. doi:10.1027/1016-9040.14.1.40

Lewis, G., Hourani, L., Tueller, S., Kizakevich, P., Bryant, S., Weimer, B., & Strange, L. (2015). Relaxation training assisted by heart rate variability biofeedback: Implication for a military predeployment stress inoculation protocol. *Psychophysiology*, *52*(9), 1167–1174. doi:10.1111/psyp.12455

Lieberman, H. R., Bathalon, G. P., Falco, C. M., Morgan, C. A., Niro, P. J., & Tharion, W. J. (2005). The fog of war: Decrements in cognitive performance and mood associated with combat-like stress. *Aviation, Space, and Environmental Medicine*, *76*(7), C7–C14.

Luque-Casado, A., Perales, J. C., Cárdenas, D., & Sanabria, D. (2016). Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological Psychology*, *113*, 83–90.

Malik, M. (1996). Heart rate variability. *Annals of Noninvasive Electrocardiology*, *1*(2), 151–181. doi:10.1111/j.1542-474X.1996.tb00275.x

Matthews, G., Campbell, S., Falconer, S., Joyner, L., Huggins, J., Gilliland, K., Grier, R., & Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*(4), 315–340. doi:10.1037/1528-3542.2.4.315

Matthews, G., & Davies, D. R. (2001). Individual differences in energetic arousal and sustained attention: A dual-task study. *Personality and Individual Differences*, *31*(4), 575–589.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S. E., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three. *Personality Psychology in Europe*, *7*, 335–350.

Matthews, G., Szalma, J., Panganiban, A. R., Neubauer, C., & Warm, J. S. (2013). Profiling task stress with the dundee stress state questionnaire. In L. Cavalcanti & S. Azevedo (Eds.), *Psychology of stress: New research* (pp. 49–90). Hauppage, NY: Nova Science.

Matthews, G., & Zeidner, M. (2012). Individual differences in attentional networks: Trait and state correlates of the ANT. *Personality and Individual Differences*, *53*(5), 574–579. doi:10.1016/j.paid.2012.04.034

McClernon, C., McCauley, M., O'Connor, P., & Warm, J. (2011). Stress training improves performance during a stressful flight. *Human Factors*, *53*(3), 207–218. doi:10.1177/ 0018720811405317

McEwen, B. S. (2001). Plasticity of the hippocampus: Adaptation to chronic stress and allostatic load. *Annals of the New York Academy of Sciences*, *933*, 265–277. doi:10.1111/j.1749-6632.2001 .tb05830.x

McEwen, B. S., & Wingfield, J. C. (2003). The concept of allostasis in biology and biomedicine. *Hormones and Behavior*, *43*(1), 2–15. doi:10.1016/S0018-506X(02)00024-7

Meichenbaum, D. (1985). *Stress inoculation training*. New York, NY: Pergamon Press.

Meichenbaum, D. (2017). *The Evolution of Cognitive Behavior Therapy: A Personal and Professional Journey with Don Meichenbaum*. New York, NY: Routledge.

Meichenbaum, D., & Cameron, R. (1989). Stress inoculation training. In D. Meichenbaum & M. Jaremko (Eds.), *Stress reduction and prevention* (pp. 115–154). New York, NY: Springer.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society*, *82*(4), 591–605. doi:10.1111/j.1469-185X.2007.00027.x

Niskanen, J. P., Tarvainen, M. P., Ranta-Aho, P. O., & Karjalainen, P. A. (2004). Software for advanced HRV analysis. *Computer Methods and Programs in Biomedicine*, *76*(1), 73–81. doi:10.1016/j.cmpb.2004.03.004

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, *33*(1), 17–33.

Orasanu, J., & Backer, P. (1996). Stress and military performance. In J. E. Driskell & E. Salas (Eds.), *Stress and human performance* (pp. 89–125). Mahwah, NJ: Lawrence Erlbaum.

Perna, F. M., Antoni, M. H., Baum, A., Gordon, P., & Schneiderman, N. (2003). Cognitive behavioral stress management effects on injury and illness among competitive athletes: A randomized clinical trial. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, *25*(1), 66–73. doi:10.1207/S15324796ABM2501_09

Porges, S. (1995). Cardiac vagal tone: A physiological index of stress. *Neuroscience & Biobehavioral Reviews*, *19*(2), 225–233. doi:10.1016/0149-7634(94)00066-A

Porges, S. (2011). *The polyvagal theory: Neurophysiological foundations of emotions, attachment, communication, and self-regulation* (Norton Series on Interpersonal Neurobiology). New York, NY: W. W. Norton & Company.

Regehr, C., Glancy, D., & Pitts, A. (2013). Interventions to reduce stress in university students: A review and meta-analysis. *Journal of Affective Disorders*, *148*(1), 1–11. doi:10.1016/j .jad.2012.11.026

Robson, S., & Manacapilli, T. (2014). *Enhancing performance under stress: Stress inoculation training for battlefield airmen*. Santa Monica, CA: The Rand Corporation.

Rose, R. D., Buckey, J. C., Zbozinek, T. D., Motivala, S. J., Glenn, D. E., Cartreine, J. A., & Craske, M. G. (2013). A randomized controlled trial of a self-guided, multimedia, stress management and resilience training program. *Behaviour Research and Therapy*, *51*(2), 106–112. doi:10.1016/j.brat.2012.11.003

Saunders, T., Driskell, J. E., Johnston, J., & Salas, E. (1996). The effect of stress inoculation training on anxiety and performance. *Journal of Occupational Health Psychology*, *1*(2), 170–186. doi:10.1037/1076-8998.1.2.170

Schwerdtfeger, A., & Derakshan, N. (2010). The time line of threat processing and vagal withdrawal in response to a self-threatening stressor in cognitive avoidant copers: Evidence for vigilance-avoidance theory. *Psychophysiology*, *47*(4), 786–795. doi:10.1111/j.1469-8986.2010.00965.x

Serino, S., Triberti, S., Villani, D., Cipresso, P., Gaggioli, A., & Riva, G. (2014). Toward a validation of cyber-interventions for stress disorders based on stress inoculation training: A systematic review. *Virtual Reality*, *18*(1), 73–87. doi:10.1007/s10055-013-0237-6

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology*, *5*, 1040. doi:10.3389/fpsyg.2014.01040

Staal, M. A. (2004). *Stress, cognition, and human performance: A literature review and conceptual framework* (NASA Tech. Memorandum 212824). Moffett Field, CA: NASA Ames Research Center.

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1228–1248. doi:10.1016/j.neubiorev.2012.02.003

Summers, R. L., Johnston, S. L., Marshburn, T. H., & Williams, D. R. (2005). Emergencies in space. *Annals of Emergency Medicine*, *46*(2), 177–184. doi:10.1016/j.annemergmed.2005.02.010

Tarvainen, M. P., & Niskanen, J. P. (2012). *Kubios HRV*. Joensuu, Finland: Biosignal Analysis and Medical Imaging Group (BSAMIG), Department of Applied Physics, University of Eastern Finland.

Thayer, J. F., Ahs, F., Fredrikson, M., Sollers, J. J., III, & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, *36*(2), 747–756.

Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, *37*(2), 141–153. doi:10.1007/s12160-009-9101-z

United States National Aeronautics and Space Administration (NASA). (2013). *International Space Station, emergency procedures 1a: Depress, fire, equipment retrieval* (No. JSC-48566). Houston, TX: NASA Johnson Space Center.

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin & Review*, *11*(2), 269–274.

Zhang, L., & Han, K. (2009). *How to analyze change from baseline: Absolute or percentage change? D-level essay in statistics*. Falun, Sweden: Dalarna University.

Tor T. Finseth is a doctoral student at Iowa State University in aerospace engineering. His research interests include the application of human factors and stress training—physiological, psychological, and human computer interaction—to improve performance on hazardous occupational tasks.

Nir Keren is an associate professor of occupational safety and a graduate faculty member at the Virtual Reality Application Center at Iowa State University. He develops virtual reality environments to study mission critical occupation employees' performance under stress.

Michael C. Dorneich is an associate professor in the Industrial and Manufacturing Engineering Department at Iowa State University. His research interests focus on creating joint human–machine systems that enable people to be effective in the complex and often stressful environments found in aviation, military, robotic, and space applications. His recent work explores the dual costs and benefits of adaptive systems, and develops approaches to mitigate shortcomings, leverage human strengths, and augment human performance for complex operational environments.

Warren D. Franke is a professor of exercise physiology in the Department of Kinesiology and a graduate faculty member in the human computer interaction program at Iowa State University. He received a PhD in applied exercise physiology from Virginia Polytechnic Institute and State University in 1991.

Clayton C. Anderson is a retired NASA astronaut. He joined NASA in 1983 and served in various roles at the Johnson Space Center. He logged 167 days in space. He completed 5 months aboard the space station in 2007 and served aboard the STS-131 crew in 2010 too. Currently he is a distinguished faculty member in the Department of Aerospace Engineering at Iowa State University.

Mack C. Shelley is a university professor of statistics and political science and in the School of Education at Iowa State University. His research, teaching, consulting, and grants activity focus on public policy and program evaluation, with emphasis on STEM education. He currently serves as chair of the Department of Political Science and formerly was the director of the Research Institute for Studies in Education at Iowa State.