EVALUATION OF BIOFEEDBACK COMPONENTS FOR THE MANAGEMENT OF ACUTE STRESS IN HEALTHCARE

Lauren R. Kennedy-Metz

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Translational Biology, Medicine, and Health

Sarah E. Parker
Nathan K. Lau
Michael A. Fox
David W. Musick

October 12, 2018
Roanoke, Virginia Tech

Keywords: biofeedback; acute stress; stress management; healthcare
EVALUATION OF BIOFEEDBACK COMPONENTS FOR THE MANAGEMENT OF ACUTE STRESS IN HEALTHCARE

Lauren R. Kennedy-Metz

Abstract

Medical error is the third leading cause of death in the United States, with surgery being a critical area for improvement. Of particular interest for this dissertation is understanding and mitigating the impact of acute stress experienced by surgeons. Previous research demonstrates the detrimental effects mismanaged acute stress can have on cognitive performance integral in optimal surgical practice. Biofeedback consists of objectively monitoring signs of stress, presenting physicians with their own physiological output in real time. Introducing appropriate, targeted coping mechanisms when they are most needed may facilitate behavioral adjustments in the face of acute stress. The goal of this dissertation research was to evaluate the potential benefit of biofeedback and coping instructions, measured by reduced perceived and physiological stress, and improved task performance. In the first study, college students participated in a first-person shooter videogame while receiving visual coping instructions. Instructions that were presented at moments of elevated stress improved downstream physiology compared to randomly administered instruction, and the presence of coping instructions was more beneficial than their absence at highly stressful times. In the second study, I adapted and validated a computer-based task to focus on components of workload experienced by physicians. This study yielded one high-stress and one-low stress version of a more demographic-appropriate task. In the final study, medical students and residents completed this task. The independent variables tested included a visual biofeedback interface, intermittent auditory coping instructions, and/or brief training on stress management and emotional intelligence. Results from this study showed that despite high cognitive workload experienced by participants receiving both biofeedback and
coping instructions, performance across stress levels was indistinguishable, and physiological indicators of stress immediately following discrete coping instructions was reflective of decreased stress. Taken together, the results of these studies validate the generation of a new lab-based task to induce stress among healthcare providers, and the physiological and performance benefits associated with physiologically-based coping instructions. Future work should investigate how these concepts can be tailored towards surgical workflow with feedback modality in mind, extended to teams, and/or scaled up to higher levels of fidelity to better capture the work environment.
EVALUATION OF BIOFEEDBACK COMPONENTS FOR THE MANAGEMENT OF ACUTE STRESS IN HEALTHCARE

Lauren R. Kennedy-Metz

General Abstract

Medical error is the third leading cause of death in the United States, with surgery being a critical area for improvement. Many medical errors are preventable, and previous research has shown that inappropriately managed acute stress is responsible for many errors. Biofeedback is one way to externalize stress states, enabling individuals to monitor their own stress, even as it is changing. With rapid advancement in technological functionality, sensors hold promise not only for personal body data, but also active interventions such as biofeedback. Biofeedback is the process of actively monitoring physiology on an external device, and updating behaviors based on that physiology. Its role as a stress management tool is growing. Commercially available sensor devices are widespread, and are generating and archiving thousands of data points every day. Rather than simply archiving this data, we can use sensor technology to inform us of our current physiological and cognitive states in real time, and use that information to alter our response to stressful stimuli to achieve more favorable outcomes. This concept can be applied specifically to address how to cope when experiencing high levels of stress. For individuals working in high-stakes environments on a daily basis, such as surgeons, using physiological data to manage stress could have the added benefit of improving performance that might otherwise suffer due to mismanaged stress. The goal of this dissertation research is to explore the potential benefit of using biofeedback and specific coping strategies to reduce stress and improve performance among healthcare providers. This research consisted of different studies, all using experimental psychology approaches and all geared towards evaluating different conditions of either visual coping instructions, training on emotional
intelligence concepts as they relate to coping under stress, visual biofeedback, and/or auditory coping instructions. The results of these studies validate the generation of a new lab-based task to induce stress among healthcare providers, support the benefit of introducing coping instructions in response to elevated physiological signs of stress, and support the need for future assessments.
Acknowledgments

I’m truly not one to subscribe to the concept of luck. I stand by the idea that “luck is the residue of good design,” a quote I display in my office or wherever I’m working to remind myself that great things don’t just happen to normal people. Great things happen when normal people work really hard, put in the time, effort, and legwork, and ultimately put themselves in the position where the outcome is nothing short of positive. To outsiders, that often appears to be luck, and even the people involved often discredit themselves by chalking it up to luck. I am passionate that this is almost always a misconception, and that luck rarely exists. What appears to be lucky is most often just the byproduct of really deliberate and thoughtful preparation.

Having said that, I do consider my overwhelmingly positive experience in graduate school to be somewhat attributable to luck. At the end of my first year my prospects were slim and the likelihood that I would find a lab and stay in the program seemed bleaker by the day. That’s when I emailed you, Sarah, and we set up a phone call. It is my opinion that you had every reason to ignore that email or politely decline. At that time, I was at the end of my rope, and any amount of kindness would have meant the world to me. But the kindness, trust, and respect you’ve given me over the years is greater than anything I could have imagined. The point is: I’m immeasurably grateful that you went out on a limb and accepted a student you hadn’t even met before, and grateful to have completed my degree with your guidance. Though you might think I’m exaggerating, you are a mentor and role model to me in many personal and professional capacities, and I couldn’t have asked for a more ideal advisor. Thank you for being open, pushing me a little further than I would have gone myself, encouraging me, and always being supportive. I’ve enjoyed learning from your example every day.
The rest of my committee has been a tremendous source of support and encouragement for me over these years as well. Thank you so much Dr. Musick, Dr. Lau, and Dr. Fox. I’m always a bit nervous standing up and sharing my work with others, but with our meetings I look forward to the challenges you all raise, thoughtful questions, and positive reception.

Thank you also to everyone I have the honor to call my mentor—namely Dr. Nichols of Roanoke College and Dr. Weiss of CRMH. I literally could not have completed my first dissertation study without the help of Dr. Nichols and the entire Psychology department at Roanoke College, or my final dissertation study without the help of Dr. Weiss and many others at the hospital. Beyond logistical support, I’ve received unending moral support and life lessons from you over the course of the years we’ve known each other, Dr. Nichols. I thought you were a huge support system during my undergraduate years, but I am so fortunate and grateful that it didn’t end there. I can’t imagine getting through graduate school without your support as well.

I’d also like to take time to acknowledge my amazing family, all who have helped me get to this point. I attribute my motivation, grit, and drive to do better to my parents, who raised me to go after my dreams, and supported me all along this (life/educational) journey. While you still may not exactly understand what I do, it means so much to me that you are interested in and care about my passion. I’m also so fortunate and grateful to have felt so much support from my brothers and sister, Adam, Austin, Alex, and Lindsey. Thanks for paving the way before me.

Thank you also to the friends from back home, and friends I’ve made along the way, whether at Roanoke College or Virginia Tech. Thank you for always putting up with my nerdiness, and encouraging it, even if inadvertently. Scotty, I know if you were still with us today you would have been supporting me all the way through this crazy experience. It was often memories of you that carried me through difficult graduate school moments.
And finally, I must acknowledge and thank my biggest support system, the one who got me through this dissertation and everything it represents—all the joys, struggles, late nights, and early mornings—my dear husband Josh, who has been by my side the entire time. Putting up with my insane schedules as I was desperately running participants at all hours, listening and paying attention as I practiced talks for the fifth time, encouraging me during the times of great uncertainty, and taking such a big risk to move to Boston at a time when we were already undergoing one million life changes have meant more to me than I can express. Thanks for being my constant companion, soundboard, and biggest support system, and for keeping me level-headed and sane.
Attribution

A brief description of the contributions made by co-authors and colleagues who assisted with the preparation of my dissertation content. The following manuscripts included in this dissertation have been previously published, or are under review, and were co-authored:


Sarah Henrickson Parker (Department of Psychology, Virginia Tech; Center for Simulation, Research and Patient Safety, Carilion Clinic; Department of Basic Science Education, Virginia Tech Carilion School of Medicine) provided assistance with the screening and selection of primary research articles included in the literature review, as well as interpretation and edits of the manuscript.


Sarah Henrickson Parker assisted with study design, data collection planning, data interpretation, and manuscript edits.


Sarah Henrickson Parker assisted with study design, data collection planning, participant recruitment, data interpretation, and manuscript edits.


Sarah Henrickson Parker assisted with study design, data collection planning, participant recruitment, data interpretation, and manuscript edits.

Patrice Weiss (Chief Medical Officer, Carilion Clinic) contributed to the development and production of a training video for the experimental protocol, as well as data interpretation and manuscript edits.
# Table of Contents

List of Figures .......................................................................................................................... xiv
List of Tables ............................................................................................................................ xv
Chapter 1: Introduction ............................................................................................................. 1
Chapter 2: Literature Review .................................................................................................... 8
Chapter 3: Timing of coping instruction presentation for real-time acute stress management: Long-term implications for improved surgical performance ......................................................... 8
Chapter 4a: Contextual inquiry and analysis ........................................................................... 10
Chapter 4b: Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task ........................................................................................................ 10
Chapter 5: Adherence to auditory coping instructions during an acutely stressful task affects stress and task performance for healthcare providers ........................................................................ 11
Chapter 2: Literature review ................................................................................................... 13
Abstract .................................................................................................................................. 13
Introduction .............................................................................................................................. 13
Materials and Methods ............................................................................................................ 22
  Search Strategy ....................................................................................................................... 22
  Inclusion/Exclusion Criteria and Quality Assessment ............................................................... 22
Results ...................................................................................................................................... 24
  Populations Studied ................................................................................................................ 26
Types of Biofeedback Used ...................................................................................................... 28
Physiological Equipment .......................................................................................................... 31
Physiological Measures and Results ......................................................................................... 33
Psychological Instruments, Measures, and Results ................................................................. 39
Performance Instruments, Measures, and Results ................................................................. 43
Features of the Intervention ..................................................................................................... 45
Discussion ................................................................................................................................. 48
  Who Benefits the Most from Biofeedback Interventions? ......................................................... 48
  Consistency and Standardization in Data Collection and Measurement ............................... 50
  Effects of Intervention Frequency and Study Duration on Collected Measures ................. 52
  Practical Implications ........................................................................................................... 54
Additional Findings and Considerations .................................................................................. 55
Conclusion ................................................................................................................................. 57
  Abstract .............................................................................................................67
  Introduction .....................................................................................................68
    Related Literature ..........................................................................................71
    Application Environment ...............................................................................74
    Current Study .................................................................................................76
  Materials and Methods ...................................................................................78
    Participants ....................................................................................................78
    Equipment ......................................................................................................79
    Procedure ......................................................................................................80
    Analysis ..........................................................................................................81
  Results ...............................................................................................................83
  Discussion .........................................................................................................90
  Conclusions .....................................................................................................97

Chapter 4a: Contextual inquiry and analysis .....................................................99
  Abstract ..........................................................................................................99
  Introduction .....................................................................................................99
  Materials and Methods ..................................................................................100
  Results .............................................................................................................102
  Discussion ......................................................................................................104

Chapter 4b: Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task ..........................................................107
  Abstract .........................................................................................................107
  Introduction ....................................................................................................108
  Materials and Methods ..................................................................................111
    Participants ....................................................................................................111
    Materials ........................................................................................................111
    Procedure ......................................................................................................112
    Analysis ..........................................................................................................113
  Results .............................................................................................................115
    Subjective Indicators ....................................................................................115
    Physiological Indicators ..............................................................................116
    Performance Indicators ................................................................................118
Chapter 5: Adherence to auditory coping instructions during an acutely stressful task affects stress and task performance for healthcare providers ................................................. 124
  Abstract .................................................................................................................. 124
  Introduction ............................................................................................................. 125
  Materials and methods ......................................................................................... 127
    Participants ........................................................................................................... 127
    Equipment ........................................................................................................... 127
    Procedure ........................................................................................................... 127
    Analysis .............................................................................................................. 131
  Results ................................................................................................................. 132
    Self-report ......................................................................................................... 132
    Physiology ......................................................................................................... 135
    Ultra-short term physiology ............................................................................. 137
    Performance ....................................................................................................... 139
  Discussion ............................................................................................................ 141
    Self-report ......................................................................................................... 142
    Physiology ......................................................................................................... 142
    Ultra-short term physiology ............................................................................. 143
    Performance ....................................................................................................... 144
  Conclusions ......................................................................................................... 147

Chapter 6: Conclusions and implications ................................................................. 149
  Timing of coping instruction presentation for real-time acute stress management: Long-term implications for improved surgical performance ........................................ 150
  Contextual Inquiry and Analysis ........................................................................ 153
  Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task 155
  Adherence to auditory coping instructions during an acutely stressful task affects stress and task performance for healthcare providers ............................................. 157
  Implications ......................................................................................................... 162
  Scientific and Practical Contributions ................................................................ 164
  Future Steps ......................................................................................................... 166
  References ......................................................................................................... 168
  Appendices ......................................................................................................... 181
    Appendix A: Publications and presentations of dissertation work .................. 181
Publications............................................................................................................................................. 181
Invited talks............................................................................................................................................ 181
Podium presentations .............................................................................................................................. 181
Poster presentations ............................................................................................................................... 182
Appendix B: Self-report instruments .................................................................................................. 183
  State Trait Anxiety Inventory for Adults, form Y-1 ........................................................................... 183
  State Trait Anxiety Inventory for Adults, form Y-2 ........................................................................... 184
  State Trait Anxiety Inventory for Adults, Short Form ....................................................................... 185
  NASA Task Load Index ..................................................................................................................... 186
Appendix C: Contextual Inquiry interview questions ........................................................................... 187
Appendix D: Heart rate variability measures ....................................................................................... 188
Appendix E: Correlations among outcome measures in Chapter 5 .................................................. 189
Appendix F: Alternative representation of Figure 2.1 ........................................................................ 191
List of Figures

Chapter 2: Literature review
Figure 2.1: Quantitative eustress-distress relationship, adapted from Allen, 1983
Figure 2.2: Example of a typical visual interface employed during HRV biofeedback training. The top panel shows sinusoidal waves associated with HR (red trace) and breathing (blue trace), demonstrating the systems’ approach to coherence as breathing rate is actively altered. The power spectrum displayed confirms that HR frequency is peaking at 0.1 Hz, the reportedly optimal breathing frequency (re-printed with permission from Lehrer et al., 2014)
Figure 2.3: Search and elimination process for systematic literature review

Chapter 3: Timing of coping instruction presentation for real-time acute stress
Figure 3.1: Change in RMSSD values surrounding peaks in HR
Figure 3.2: Change in RMSSD values surrounding stimulus presentation
Figure 3.3: Time in seconds between kills across conditions
Figure 3.4: Time in seconds between deaths across conditions

Chapter 4b: Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task
Figure 4b.1: MATB-II interface
Figure 4b.2: Mean perceived cognitive workload across different versions of MFMG task
Figure 4b.3: Mean normalized RMSSD and pNN50 values across different versions of MFMG task
Figure 4b.4: Mean normalized SDNN values across different versions of MFMG task
Figure 4b.5: Mean performance scores across different versions of MFMG task

Chapter 5: Adherence to auditory coping instructions during an acutely stressful task affects stress and task performance for healthcare providers
Figure 5.1: Experimental set-up
Figure 5.2: Self-reported stress state comparing high-stress and low-stress conditions
Figure 5.3: Self-reported cognitive workload state comparing high-stress and low-stress conditions
Figure 5.4: Self-reported cognitive workload state across stress levels and group assignment
Figure 5.5: Time-domain measures of HRV comparing high-stress and low-stress conditions
Figure 5.6: Frequency-domain measures of HRV comparing high-stress and low-stress conditions
Figure 5.7: Ultra-short term SDNN values comparing 1) 10-second, b) 20-second, and c) 30-second time windows before and after coping instruction adherence
Figure 5.8: Overall MATB-II performance comparing high-stress and low-stress conditions
Figure 5.9: Overall MATB-II performance across stress levels and group assignment
List of Tables

Chapter 2: Literature review
Table 2.1: Master table with comprehensive study details
Table 2.2: Study samples across studies included in systematic literature review
Table 2.3: Biofeedback modality across studies included in systematic literature review
Table 2.4: Physiological measures, equipment, and results across studies included in systematic literature review
Table 2.5: Description of components within physiological measures and psychological assessments across studies included in systematic literature review with significant changes reported
Table 2.6: Psychological measures, instruments, and results across studies included in systematic literature review
Table 2.7: Performance measures, instruments, and results across studies included in systematic literature review
Table 2.8: Intervention and study duration and frequency across studies included in systematic literature review

Chapter 3: Timing of coping instruction presentation for real-time acute stress management: Long-term implications for improved surgical performance
Table 3.1: Sample sizes, mean values, p values, and standard deviations of participants’ overall state of stress and perceived cognitive workload
Table 3.2: Mean RMSSD values surrounding peaks in HR and stimulus presentation across conditions
Table 3.3: Sample sizes, mean values, and standard deviations of participants’ RMSSD values surrounding peaks in HR and stimulus presentation
Table 3.4: Mean time between kills and deaths across conditions
Table 3.5: Sample sizes, mean values, and standard deviations of participants’ performance metrics, including time between kills and time between deaths across conditions

Chapter 4a: Contextual inquiry and analysis
Table 4a.1: Interview questions and follow-up questions

Chapter 4b: Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task
Table 4b.1: Characteristics for each subtask of MFMG

Chapter 5: Adherence to auditory coping instructions during an acutely stressful task affects stress and task performance for healthcare providers
Table 5.1: Experimental groups

Chapter 6: Conclusions and Implications
Table 6.1: Summary of overall results
Chapter 1
Introduction

The overall goal of this dissertation was to determine if explicit feedback, in the form of heart rate biofeedback and/or coping instruction presentation, could be used as a stress management tool to decrease self-reported and physiological signs of stress, and increase task performance on stress-inducing, computer-based tasks. My work was done within a translational framework by utilizing experimental methods and approaches to contribute to the solution of a clinical problem. In this chapter, I will briefly discuss the context in which this question arose, the impact of acute stress on surgical performance, the potential implications of biofeedback as a stress management intervention, and an introduction to the thesis work that follows. The research question that guided this work is: Can different components and combinations of a biofeedback intervention (heart rate biofeedback, explicit coping instructions, emotional intelligence training) reduce stress and improve performance among a population of healthcare practitioners? Overall, I hypothesized that high levels of appropriately designed cognitive support (i.e. some format of biofeedback combined with some format of directive coping instructions) would effectively reduce stress and improve performance.

Surgery is a compelling context in which to study the effects of stress on performance, because there is the potential for poor performance to manifest as patient harm. Data shows that the most recent estimates of surgical outcomes place medical error as the third leading cause of death in the United States, after only heart disease and cancer (Makary & Daniel, 2016). However alarming that figure may seem, many have argued that this is a conservative estimate. Despite the traditional view of surgical quality and safety, which posits that outcomes are based on the interactions between the clinicians’ technical skill and the patient risk factors, a more
modern framework introduced the systems approach to surgical quality, incorporating additional features influencing surgical practice into the equation (Vincent, Moorthy, Sarker, Chang, & Darzi, 2004).

Systems factors, including a range of non-technical skills and cognitive influences such as stress, have been demonstrated to drastically affect surgical procedures (Anton, Montero, Howley, Brown, & Stefanidis, 2015). Of those factors, acute stress is of particular interest due to its inextricable influence on an individual’s physiological response, cognitive processing, and subsequent behavioral reactions.

Stress, and more specifically acute stress, is a transient and dynamic response internally initiated in response to environmental demands, and mediated by perceptions of available resources compared to perceptions those environmental demands. The framework of cognitive appraisal distinguishes between challenging and threatening stressors. This theory posits that when available resources are perceived to be sufficient to meet the demands of the environment, we feel challenged and are motivated to take on the stressor. However, when the available resources are insufficient to meet the demands posed by the environment, the result is a feeling of distress and the appraisal of a threat (Lazarus & Folkman, 1984).

When faced with a threatening stressor, our physiological profiles adaptively respond by initiating a series of endocrine and hormonal changes designed to protect us from eminent danger (Tanev, Saadi, Hoppe, & Sorensen, 2014). This response is often referred to as the “fight-or-flight” response, and is characterized by an increase in heart rate, galvanic skin response, perspiration, and pupil dilation; a change in breathing rate; and decrease in salivary production, peripheral blood flow, and a corresponding decrease in distal skin temperature, among other changes (Venables & Fairclough, 2004; Zhai & Barreto, 2006). These changes, associated with
the autonomic nervous system, are known to be vagally-mediated (Thayer, Hansen, Saus-Rose, & Johnsen, 2009). As this response continues, systemic cortisol release initiates as well (McEwen, 2006).

Once systemic cortisol levels reach neural areas, cognitive functioning underlying key areas begins to suffer (McEwen, 2006). Flooding of cortisol primarily targets the prefrontal cortex, the amygdala, and the hippocampus, thereby affecting executive functioning and attentional processes. Some of the processes known to be disrupted as a result include decision-making, working memory, and attention functions (LeBlanc, 2009). When cognitive performance across these domains is threatened, there is also the potential for systems factors to be affected as well, including social processes, communication, teamwork, and situational awareness. Theories of cognitive capacities, however, can introduce potential solutions to ameliorate potential negative side effects associated with excessive stress.

Cognitive Load Theory (CLT) is one example of a cognitive architecture that takes into account the limited processing capabilities associated with demanding tasks. CLT rests on the assumption that our finite processing capacity can be manipulated according to three categories of mental load: intrinsic, extraneous, and germane cognitive load (Van Merriënboer & Sweller, 2010). That which is most pertinent to the work described in this dissertation is germane load, defined as the load induced by the conscious effort of applying strategies to increase task efficiency (Schnitz & Kürschner, 2007). Refer to Chapter 3 for a more detailed explanation of CLT, its three categories of mental load, and their relationships with variables studied in this dissertation.

Importantly, the cognitive changes characterized by excessive levels of acute stress begin with the aforementioned basic physiological changes. Each of these basic physiological changes
associated with the acute stress, “fight-or-flight” response is detectable via noninvasive methods (or, in the case of cortisol via minimally invasive methods) (Zhai & Barreto, 2006). A comprehensive theory outlining the integration of physiological and cognitive changes experienced under acute stress, as well as how those changes affect performance outcomes, can be found in the literature surrounding the neurovisceral integration model (Thayer, Åhs, Fredrikson, Sollers, & Wager, 2012; Thayer et al., 2009). This model outlines central nervous system involvement in key cognitive resources and subsequent behavioral adaptations. According to the model, a broad network of associated neural structures, which are affected by changes in the cardiovascular system, enables flexibility and adaptability in the face of environmental demands. Maladaptive physiological and cognitive patterns as well as behaviors emerge when this network is uncoupled and not working appropriately.

Furthermore, each of the physiological changes encountered under stress is detectable, and able to be calculated upon in real time, via sensor technology. Sensors afford the ability to continuously and passively collect thousands of data points, which, when processed appropriately, can provide near-real time estimates of underlying cognitive states. With sensors, these estimates can be continuously updated in a very sensitive and reliable way, enabling the capture of changes over time. An extension of this functionality is the ability to visualize and monitor these changes as they are occurring.

Visualizing and monitoring physiological changes in real time, and using that information to inform subsequent behaviors, is the backbone of biofeedback (Schwartz, 2010). Biofeedback as a concept and technological system has existed for decades, but with more recent and rapid advancements in sensor technology, it has become extremely robust and sensitive (Al Osman, Eid, & Saddik., 2013). The computationally robust methodology, unobtrusive equipment, and
information display opportunities associated with biofeedback make it an ideal mechanism for stress management training within the context of healthcare.

In Chapter 2, I provide an in-depth systematic review that conveys additional evidence that biofeedback interventions are most effective when they are tailored to the context, work, and demographic (Kennedy & Parker, 2018b). In surgery specifically, research has shown that psychomotor performance suffers as a result of acute stress, affecting economy of motion and number of errors (Arora, Sevdalis, Aggarwal, et al., 2010). Excessive stress also leads to impaired decision-making, affecting surgical performance (Arora, Sevdalis, Nestel, et al., 2010). In novice surgeons particularly, alertness, concentration, focus, and efficiency of action improve with moderate levels of intraoperative stress, but are all compromised when stress (Hassan et al., 2006) and cortisol (Arora, Sevdalis, Aggarwal, et al., 2010) are too high. Training to manage stress within an optimal range could appropriately address adverse physical and cognitive reactions.

Though biofeedback has not been evaluated in the healthcare domain for its utility in stress management, evidence does support the efficacy of integrating information displays into clinical practice (Yang, Kang, & Lee, 2015). There is potential for an intervention such as biofeedback to induce further cognitive load, by way of increasing the level of germane load required to consciously improve task performance (Schnotz & Kürschner, 2007). Ensuring the effectiveness of a biofeedback intervention for stress management in healthcare thus requires establishing optimal usability of the technical system. This is critical because an inappropriate design could lead to interruptions, distracting information, and ambiguous coping instructions (Kontogiannis & Kossiavelou, 1999). Design components must maximize stress management
and minimize adverse consequences. To capitalize on natural human processing abilities, the integration of components must be examined.

Due to the already complex socio-technical system represented by healthcare, the introduction of new information or displays must be carefully considered and evaluated. Ecological interface design is a theoretical framework that incorporates the interactions between visual processing and higher level cognitive control in an effort to support cognition (Vicente & Rasmussen, 1992). This framework, along with others, supports the use of human-computer interfaces or displays to externally represent information that might otherwise be delegated to working memory (Hegarty, 2011). With this in mind, however, we must also acknowledge that presenting the same information in different formats can lead to significant differences in cognition and subsequently in performance (Breslow, Trafton, & Ratwani, 2009; Hegarty, 2011; Novick & Catley, 2007; Peebles & Cheng, 2003; Sanfey & Hastie, 1998; Simkin & Hastie, 1987). Based on the context being explored, arriving at the appropriate format for information display may be more critical for certain variables.

For example, given the time-sensitivity of the work domain and the acute stress response, presentation timing is a crucial variable to address in the realm of stress management in healthcare (Vicente & Rasmussen, 1992), work that will be described in Chapter 3, as well as Chapter 5. Chapter 3 describes my initial exploration into the concept of incorporating coping instructions amid an acutely stressful task. My key focus in this project was in determining the ideal timing of coping instruction presentation. A convenience sample of undergraduate psychology and biology students was recruited for this study, and the task involved completing a cognitively demanding, dynamic visual-spatial first-person shooter videogame. Visual coping instructions were displayed on various time schedules, but the presentation timing of most
interest administered coping instructions in response to elevated physiological signals indicative of high stress (Kennedy & Parker, 2018b). We find support for these concepts from other researchers, who have shown that providing psychophysiodriven adaptive aiding during complex aviation tasks enhances performance (G. F. Wilson & Russell, 2007).

Following this study, the next challenge was in translating these concepts to healthcare practitioners and involving a healthcare-oriented task. In Chapter 4a and 4b, I provide insight into the process to ensure an appropriate experimental design and task. Chapter 4a provides details surrounding interviews conducted with target end users, attending surgeons, to determine the following: which design components would be considering integral, disruptive, and/or ambiguous; how to overcome some of the shortcomings observed in the study described in Chapter 3; what the appropriate application of this intervention would be; and what the overall interest in the development of a biofeedback intervention would be. These interviews were meant to be informational, and were not deliberately powered for statistical findings.

With these design and application implications in mind, the next step was to determine the appropriate stress-inducing task for a recruitment sample of medical students and surgical trainees. To obtain maximum experimental control, this required relying on lab-based approaches. Chapter 4b describes the development and validation of a computer-based task, Medically-Focused Multitasking Game, adapted from a previously existing platform for a population of healthcare specialists (Kennedy & Parker, 2017). These two pilot studies, though underpowered in some ways, directly influenced the experimental design of the study that followed, described in Chapter 5.

Equipped with an appropriate, highly controlled task and experimental setting, along with sensitive and reliable outcome measures, the final component of my systematic evaluation was in
the integration of the main components of biofeedback and coping instructions in one experimental design. The details of this phase of the dissertation work are reported in Chapter 5. The work completed towards my dissertation is translational in nature, with methodological and analytical focus on fine-grained physiological changes ultimately contributing to the goal of improving clinical care.

The overarching research question guiding all of the work presented in this dissertation is: Can different components and combinations of a biofeedback intervention reduce stress and improve performance among a population of healthcare practitioners? A literature review and 3 major studies are reported here. Each contribute to the overarching research question, and were additive in their contribution to answering the research question:

Chapter 2: Literature Review

Objective

- To evaluate the evidence of how biofeedback is currently being used to address stress management for the purposes of enhancing performance.

Chapter 3: Timing of coping instruction presentation for real-time acute stress management: Long-term implications for improved surgical performance

Research Questions and Hypotheses

1. Do the perceived levels of stress and cognitive workload change when the task difficulty increases?
Hypothesis 1: Responses to the self-report questionnaires will indicate a higher level of stress and a higher perceived cognitive workload in the more difficult conditions, indicative of a higher level of intrinsic load, with increased task difficulty.

2. When heart rate is substantially elevated, is there a difference between receiving and following a coping instruction at this time, versus not receiving a coping instruction at this time?

Hypothesis 2: Physiological measures downstream of times of substantially elevated heart rate will be representative of a lower level of stress when a coping instruction is presented and followed, compared to when a coping instruction is not presented or not followed. Removing the requirement to devote cognitive resources to coping due to this intervention will decrease the level of germane load.

3. Does receiving and following a coping instruction have a different effect when the instruction is physiologically relevant, versus when it is physiologically arbitrary?

Hypothesis 3: Physiological measures downstream of times when coping instructions are presented will reveal that stress is more effectively reduced when the coping instructions presented are based on changes in physiology, compared to when coping instructions presented are not based on changes in physiology. Providing an explicit coping strategy when it is needed will be less distracting than providing one when it is not needed, and will decrease the level of germane load.

4. Does performance differ as a result of increased intervention adherence?

Hypothesis 4: Performance will improve when the intervention is adhered to more consistently over the course of the condition and task. Following the instructions of
an explicit coping strategy throughout the task will reduce the amount of germane load required.

Chapter 4a: Contextual inquiry and analysis

Objectives

- To gauge whether a biofeedback intervention would be welcomed by target end users themselves, surgeons.
- To determine appropriate design considerations, such as specific feedback components, their format, and their modality.

Chapter 4b: Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task

Research Questions and Hypotheses

1. Do the perceived levels of stress and cognitive workload change when the task difficulty increases?
   
   Hypothesis 1: Subjective perceptions of stress and cognitive workload will increase as the task difficulty advances from the least difficult to the most difficult conditions.

2. Do physiological indicators of stress, derived from electrocardiography, change when the task difficulty increases?
   
   Hypothesis 2: Objective physiological indicators of stress will increase as the task difficulty advances from the least difficult to the most difficult conditions.

3. Does overall task performance change when the task difficulty increases?
Hypothesis 3: Overall performance on the task will decrease as the task difficulty advances from the least difficult to the most difficult conditions.

4. Do these three categories of dependent variables follow the same pattern of change in response to increasing task difficulty?

Hypothesis 4: Subjective, objective, and performance results will align and thereby indicate which condition is considered the least stressful and which is considered the most stressful.

Chapter 5: Adherence to auditory coping instructions during an acutely stressful task affects stress and task performance for healthcare providers

Research Questions and Hypotheses

1. Do perceptions of stress and cognitive workload differ as a result of different intervention components?

Hypothesis 1: Subjective reports of stress and cognitive workload will be lower when participants receive the highest degree of real-time cognitive aid (Group 3; biofeedback, coping instructions).

2. Do physiological measures of stress differ as a result of different intervention components?

Hypothesis 2: Objective indicators of stress will be lower when participants receive the highest degree of real-time cognitive aid.

Hypothesis 3: Ultra short-term objective indicators of stress will be lower immediately following adherence to discrete intervention events (coping instructions).
3. Does performance differ as a result of different intervention components?

Hypothesis 4: Task performance will be higher when participants receive the highest degree of real-time cognitive aid.

The research described here contributes to the study of biofeedback by incorporating healthcare professionals into the scope of its application. It broadens the utility of biofeedback for stress management purposes to a novel demographic, and expands its functionality to additionally include physiologically-based coping instructions. In the realm of biofeedback, this can be considered a substantial step, but the implications for this work are far more substantial when considering the potential influence on healthcare training and practice. Addressing the detrimental effects of acute stress in healthcare, and adding to the literature surrounding how to actively intervene with this process is critical. This work supports the use of unobtrusive physiological sensors to capture data indicative of stress states in real time, and to go one step further by using this information to alter behavior. Behavioral adaptation in the context and setting of the operating room has implications for surgical performance, and more importantly, for patient safety. In the realm of the translational spectrum, the basic science work described contributes to the evidence-based foundation required to systematically implement and evaluate future interventions in applied and meaningful ways.
Chapter 2
Biofeedback as a stress management tool: A systematic review

Abstract

Inappropriate management of acute stress can negatively affect cognition and task performance. Frequently occurring acute stress encounters can lead to cardiovascular and immunity deficiencies, and psychological disorders such as depression, fatigue, and burnout. Biofeedback can be used as a non-invasive, passive, continuous method of managing stress in real time. A systematic review of biofeedback as a real-time stress management intervention for non-patients was conducted to identify literature between 2000 and 2017, yielding 17 studies evaluating physiological, psychological, and/or performance metrics. Participants represent convenience samples (N=9 studies) and deliberately selected samples whose optimal performance under stress is critical for occupational success (N=8 studies). Various methods to collect data, display biofeedback, induce stress, and measure performance were reported. Overall, biofeedback is an effective intervention that can be used to reduce physiological and subjective stress, and enhance performance. This is especially true among professionals whose job performance requires appropriate stress management.

Introduction

Imagine being a volunteer at the Boston Marathon in 2013, and watching as the celebratory scene transforms into shock and chaos. The sounds of explosions fill your ears. People scatter to determine the cause of chaos and help those in need. Your palms start sweating, your pupils dilate, and your heart rate and breathing rate increase. Your body is preparing you to
escape to safety as efficiently as possible. Jumbled thoughts ensue, but your concern is physically leaving danger to find safety. This is an illustration of the adaptive and beneficial effects of the short-term acute stress response initiated by the sympathetic nervous system (Allen, 1983; Greenberg et al., 1980; Venables et al., 2004; Zhai et al., 2006).

Now imagine being a physician in one of Boston’s hospitals, suddenly flooded with patients in need of urgent care, with no prior knowledge of the situation and no time to waste. Identical physiological cascades take place by way of sympathetic nervous system activation, characterizing the acute stress response. Acute stress is defined by its short-term, immediate, and transient physiological changes that occur when we are confronted with some form of perceived danger (Tanev et al., 2014). But the chemical and endocrine imbalances intended to prepare the body to fight or flee advance over minutes as systemic cortisol release reaches the hippocampus, amygdala, and prefrontal cortex (LeBlanc, 2009; McEwen, 2006). As a victim attempting to escape to safety, the cognitive side effects of this acute stress endocrine cascade—including decision-making, attentional, and memory deficits—have little effect on the overall goal. But to the physician, whose performance relies on optimal functioning of decision-making, memory recall, and cognitive control and determines the outcome of the patient, the cognitive alterations associated with cortisol release stimulated by an acute stressor are detrimental and need to be managed appropriately.

In addition to the detrimental effect of discrete stressors on cognition and performance, repeated exposure to stressors with little recovery time between events also presents a danger. Excessive levels and extended durations of acute stresses characterize chronic stress (McEwen, 2006; Tanev et al., 2014). Burnout is a subset of prolonged or chronic stress that is extremely prevalent among helping professions, particularly among healthcare practitioners (Balch et al.,
2009; Rosenstein, 2012). Physician burnout is associated with a threat to patient safety, diminished quality of patient care, and an increase in medical errors and adverse events (Balch et al., 2009).

Managing stress within an optimal range and in moderation ensures the highest potential for peak performance (Allen, 1983; Hupbach et al., 2012; Vine et al., 2013). The quantification of positive stress, or eustress, and detrimental stress, or distress, can be overlaid with the Yerkes-Dodson law to illustrate the curvilinear relationship between performance and stress (Figure 2.1; see Appendix F for an alternative representation). The amount of stress required to achieve optimal performance varies by activity and individual, emphasizing the need to personalize stress management approaches.

![Figure 2.1 Quantitative eustress-distress relationship, adapted from Allen, 1983](image)

The stress referred to in Figure 2.1 can be physical, mental, or emotional (Greenberg et al., 1980). Emotional and cognitive/mental stress are of particular interest because they specifically contribute to disruptions in decision-making, attentional, and memory processes (Cohen et al., 2015). Furthermore, cognitive stress can often have a more direct impact on performance (Sime, 2007). In high risk professions, poor performance is unacceptable. High-risk professions are characterized by frequent occurrences of high-workload situations managed by small groups operating within risky environments (Klampfer et al., 2001). Examples commonly include teams found in the military, air craft crews, surgical teams, police force, teams working
in nuclear power plants, etc. When performance is affected, particularly in such high-risk professions where physically and cognitively demanding acute stressors are commonplace and any error can have catastrophic consequences, understanding the contributors to performance under stress and developing mitigation strategies is necessary (Delahaij et al., 2017; Delahaij et al., 2011).

A comprehensive understanding of the interactions among physiological, cognitive, and performance variables occurring under stress has been outlined in the neurovisceral integration model (Thayer et al., 2012; Thayer et al., 2009). This model is built on the basic empirical evidence that cardiovascular functioning, indexed by heart rate (HR) and heart rate variability (HRV), is mediated via dual-innervation of both the sympathetic nervous system and the parasympathetic nervous system. Though both branches typically achieve a dynamic balance in the resting state, humans become susceptible to physiological and cognitive disruptions—starting at the level of the prefrontal cortex (PFC) and ultimately influencing HRV and cognitive performance—when a static imbalance is reached. In the context of acute stress, static imbalance is initiated as a physiologically-adaptive response to environmental demands, with potentially maladaptive cognitive consequences in the absence of appropriate behavioral adaptation. The neurovisceral integration model proposes that an overarching central nervous system (CNS) network controls psychophysiological attentional, emotional, and executive functioning resources. According to the model, associated neural structures—including the PFC, cingulate cortex, insula, amygdala, hypothalamus, and others—are responsible for coordination of behavioral adaptations. Thus, this broad and interactive CNS network enables flexibility and adaptability in the face of environmental demands. Maladaptive patterns and behaviors emerge when this network is uncoupled and not working appropriately (Thayer et al., 2012, 2009).
The neurovisceral integration model, along with empirical pharmacological and neuroimaging evidence (Thayer et al., 2012), supports evidence that cortical activity (namely PFC cortical activity) modulates associated cardiovascular function (namely HRV). The model further provides empirical evidence that performance on executive functioning (i.e. decision making and memory processes) and attention tasks is affected by the same top-down process. Though stress can impair these executive functioning and attentional abilities and can negatively affect HRV, increasing resting levels of HRV as a means of stress management can improve cognitive functioning (Thayer et al., 2009). This can be attributed to the role of HRV in performance on tasks recruiting PFC-localized abilities (i.e. executive functioning and attention processes). Therefore theory suggests, and evidence confirms, that enhancing HRV can improve PFC-localized cognitive abilities and in turn, improve performance on tasks relying on those abilities. In most experimental literature, HRV is measured as a dependent variable, but the neurovisceral integration model incorporates HRV as an independent variable, capable of active manipulation to produce a desired outcome or set of outcomes (Thayer et al., 2012, 2009). Introducing behavioral programs has been demonstrated to be an effective way to improve cognitive functioning and performance by way of manipulating HRV (Thayer et al., 2009). One such behavioral intervention capable of directly addressing HRV is biofeedback.

Biofeedback is a mind-body intervention that has the benefit of being unobtrusive, passive, and continuous. Unlike many mind-body interventions addressing stress (e.g. Lehrer, 2007), biofeedback externalizes an individual’s physiological state and allows the user to monitor changes in real time (Schwartz, 2010). Thus, as the stress response unfolds, users can track its progression, and take steps to preemptively mitigate any detrimental changes. With training over time, these processes can become more automatic, with little or no reliance on the
biofeedback instrument (Schwartz, 2010). While biofeedback can measure a variety of psychophysiological variables and can deliver data back to the user through a variety of modalities, HRV biofeedback as a specific subtype of biofeedback is a standardized method with manuals, protocols, and guides to illustrate its uses (Eddie et al., 2015; Lehrer et al., 2013; Lehrer et al., 2000).

HRV biofeedback consists of monitoring HR oscillations using a visual display, and learning to voluntarily generate a simple sinusoidal curve on the display, representing cross-coherence, by actively altering one’s breathing rate. Cross-coherence between systems considered to be oscillatory—those with repetitive temporal variation—occurs when two rhythmic systems, such as heart rate and breathing rate, match in frequency (McCraty et al., 2009). By changing the frequency of breaths, a HRV biofeedback user can induce HRV changes, blood pressure changes, activate the baroreflex, and achieve coherence among the cardiovascular systems. Breathing at a rate of 9 to 24 breaths per minute, which translates to 0.15 to 0.4 Hz in frequency, is typical for most individuals, though this rate and its inverse, frequency, varies depending on the individual and the context. HR oscillations that correspond to this breathing rate are referred to as respiratory sinus arrhythmia (RSA), which has the same range in frequency (0.15 to 0.4 Hz) as the “high frequency” (HF) band of the HRV power spectrum. The HF/RSA index is associated with regulatory functions and parasympathetic tone via vagal nerve innervation of the heart (Lehrer et al., 2014).

The goal in HRV biofeedback training is for the user to arrive at the appropriate breathing rate that coincides with their RSA frequency, resulting in a 0⁰ phase relationship, or optimal coherence/overlap, of the two sinusoidal curves representing 1. breathing rate, and 2. HF/RSA amplitudes (Figure 2.2). Achieving this goal results in a coherent HRV waveform,
where both traces are matched over time (refer to McCraty et al., 2009 for details on calculating the heart rhythm coherence ratio). In practice, this $0^\circ$ phase relationship, or ideal HRV waveform, is observed as times when heart rate and breathing occur concurrently. In other words, optimal coherence is obtained when heart rate rises simultaneously as a user inhales, and falls simultaneously as a user exhales. Additional influences on overall coherence and a more stable physical state include the phase relationship between HR and blood pressure, as well as the baroreflex, which is a blood pressure sensor and modulator (Lehrer, 2013b). Taken together, evidence suggests that a breathing frequency of 0.1, or a breathing rate of 6 breaths per minute, typically results in maximum benefit (Vaschillo et al., 2006).

![Figure 2.2 Example of a typical visual interface employed during HRV biofeedback training. The top panel shows sinusoidal waves associated with HR (red trace) and breathing (blue trace), demonstrating the systems’ approach to coherence as breathing rate is actively altered. The power spectrum displayed confirms that HR frequency is peaking at 0.1 Hz, the reportedly optimal breathing frequency (re-printed with permission from Lehrer, et al., 2014)](image-url)

Achieving coherence using HRV biofeedback techniques requires the appropriate equipment, software, and training, but this can typically be obtained within a few minutes for
naïve users (Lehrer et al., 2014; Vaschillo et al., 2006). Apart from HRV biofeedback training, other biofeedback approaches can incorporate minimal or no training, depending on how intuitive the format of feedback is. Using biofeedback and self-tracking of physiology to manage symptoms of acute and chronic stress to improve health and performance is becoming ubiquitous due to the affordability, sensitivity, and functionality of such devices in the form of high-grade consumer products. In high-risk settings, this type of passive monitoring allows for continuous data collection, without requiring a practitioner to stop a critical task, making it a highly valuable potential mechanism for performance feedback (Kennedy-Metz et al., under review).

The effectiveness of biofeedback has been documented in addressing psychiatric (Schoenberg et al., 2014) and physical (Khazan, 2013) disorders, improving unhealthy symptoms and performance (Gevirtz, 2013). This can partially be explained by operant conditioning mechanisms occurring during breathing-based biofeedback interventions such as HRV biofeedback. Notably, evidence has shown that volitional breathing behaviors can be modified via operant conditioning procedures such as biofeedback (Ley, 1999). However, the effectiveness of biofeedback in enhancing health by managing one’s stress levels, especially among non-patient populations, has received less attention, but is promising in concept (Lehrer, 2007; Lehrer et al., 2008). There are a number of professionals in high-risk occupations whose performance depends on appropriate management of stress under pressure, and whose inappropriate management of stress can lead to dire consequences. These professionals, including physicians, police officers, and soldiers, represent individuals who stand to benefit the most from proper stress management (Frazier et al., 2018). Given that these individuals more frequently experience high-level acute stressors, we would expect to see the greatest benefit as a result of a stress reduction intervention among these populations.
To better understand the benefits of stress reduction techniques in high-risk professions, it is imperative to identify appropriate measures, and particularly the consistency of these measures across studies. Stress reduction techniques such as progressive muscle relaxation, different forms of meditation, cognitive-behavioral therapy, hypnosis, guided imagery, and biofeedback have been evaluated in work settings (Bormann et al., 2006), biofeedback has the advantage of providing objective guidance in a passive manner, thus incurring minimal interruptions into daily workflow. The effectiveness of stress reduction techniques such as biofeedback can readily be evaluated through a variety of modalities, but if physiological, psychological, and/or performance indices and assessments are not consistent across studies, the ultimate implication is a lack of comparability. At the same time, standardizing performance assessment across all populations could contribute to a lack of sensitivity to specific contexts and task requirements expected of those populations. Using lab-based assessments primarily relied upon when recruiting convenience samples, for example, reduces the ecological validity when applied to deliberately recruited populations.

The goal of this systematic literature review is to evaluate the evidence of how biofeedback is currently being used to address stress management for the purposes of enhancing performance. Specifically, I am asking the following questions:

1. How does biofeedback affect physiological, psychological, and performance measures with respect to the type of population recruited? Which populations benefit the most from biofeedback interventions?

2. How consistent or standardized are the approaches used to answer these questions (including physiological indices and equipment, psychological constructs and instruments, and performance tasks and assessments)?
3. How do physiological, psychological, and performance outcomes change with respect to the frequency of the intervention and the duration of the study?

Our scope includes all populations of healthy individuals across any profession, with a special interest in high-risk professions. In the following description, I synthesize the literature to identify patterns and determine an effective approach based on the evidence available.

Materials and Methods

Search Strategy

Articles were originally selected by searching EBSCOhost, JSTOR, PsycINFO, PubMed, and others using the university library search features. Search terms included (real-time AND biofeedback AND stress AND (professional OR practitioner)), and articles were included between January 1, 2000 and June 1, 2017. One researcher (LK) completed this step, ensuring that search terms scanned all available text, rather than searching within titles or keywords only. Additional filters were imposed to narrow the search to articles written in English and excluded magazine, trade publication, and newspaper articles. Following this compilation, duplicates acquired through more than one source were removed by one researcher (LK).

Inclusion/Exclusion Criteria and Quality Assessment

Abstracts were then reviewed according to pre-defined inclusion and exclusion criteria by one researcher (LK). Inclusion criteria ensured that resulting articles were limited to original research articles and included biofeedback administered during the task, used as the primary experimental procedure, and for stress management purposes explicitly. Exclusion criteria eliminated articles consisting of non-human and non-trial articles, as well as articles evaluating
patient populations. Types of articles that were removed in this step included methods papers, reviews, instrument design articles, protocols, animal studies, machine uses, molecular approaches, case studies, and abstracts from conference proceedings.

All remaining articles underwent a quality assessment by one researcher (LK) to ensure rigorous methods and transparent reporting. Final articles were considered high quality and included in the review only if they described the biofeedback modality, participant/sample demographics, and intervention design and conditions. Final articles also had to report two of the following three data types: physiological, psychological, and performance measures. The final articles included in the review were approved independently by two researchers (LK, SHP) to ensure agreement (Figure 2.3).

**Figure 2.3** Search and elimination process
Results

Of the final 22 articles, the three articles written by Prinsloo and colleagues (Prinsloo et al., 2011; Prinsloo, Derman, Lambert, & Rauch, 2013; Prinsloo, Rauch, Karpul, & Derman, 2013) report different sets of results from the same study. For the purposes of accurate evaluation, the results of these three articles are considered to represent one study (N of articles=22; N of studies=20).

The 22 articles (20 studies) remaining almost exclusively evaluate and address mental, cognitive, or emotional stress. There was one exception, which evaluated and addressed the management of physical stressors among college students (Evetovich, Conley, Todd, Rogers, & Stone, 2007). Since this is the only article of its kind dealing with physical stress, I will report the results separately to exclusively capture mental stress in the following results. Evetovich and colleagues (2007) recruited 12 college students to evaluate the effectiveness of visual mechanomyographic biofeedback on forearm muscle flexion and relaxation. Performance was measured by calculating the number of repetitions of the seated preacher curl exercise, while physiological signals were measured using BioPac acquisition hardware and software to capture mechanomyographic and electromyographic amplitudes. This study took place over 9 days, and each participant completed 5 sessions total. Performance analysis revealed no significant differences between the experimental and control groups, while physiological analysis revealed a decrease in both mechanomyographic and electromyographic amplitudes in the experimental versus control groups.

Two of the 22 articles (20 studies) reported correlational results only, eliminating the ability to detect and report causal relationships between the biofeedback intervention and outcome measures (van Dijk, Westerink, Beute, & IJsselsteijn, 2015; Zauszniewski, Au, &
Musil, 2013). Because my focus is on hypothesis-driven, cause-and-effect results of introducing biofeedback, I will report these results separately. Van Dijk and colleagues (2015) recruited 74 healthy adults as participants to determine the synchronicity between physiological indicators of stress and self-perceived stress during and following a biofeedback intervention. Participants in the experimental groups completed a variety of short, computer-based, stress-inducing tasks while viewing their trending HR from the preceding three minutes, along with their HR value displayed in beats per minute (BPM). One experimental group was told that the display represented their current HR while the other experimental group was informed that the display represented their current stress level. The control group completed the same tasks but did not receive physiological feedback. Analysis of physiological data revealed that for the group that was viewing their “current HR” in real time, there was a higher correlation between actual HR and their self-reported momentary and retrospective stress compared to the control group. Results of psychological assessments revealed that there was a positive correlation between the personality factor of neuroticism and momentary and retrospective stress in the control group. Furthermore, there was a positive correlation between the personality factor of anxiety sensitivity and momentary and retrospective stress in the experimental group viewing their “current stress” in real time. Their results broadly support the perspective that monitoring physiological signals may enhance body awareness, and that distinct personality factors may distinguish those individuals who benefit from biofeedback and those who may not.

Zauszniewski and colleagues (2013) recruited 20 grandmothers raising their grandchildren in an effort to reveal the relationships between physiological indicators of stress and self-perceived stress as a result of a biofeedback intervention. Participants used the handheld StressEraser to view their HRV waveform in real time with the goal of achieving coherence. A
higher coherence index is known to be representative of a lower state of stress (Childre & McCraty, 2010). Overall, analysis revealed a negative correlation between perceived stress and coherence, negative emotions and coherence, and depressive cognitions and coherence. These correlations indicate that as coherence decreases—a physiological indication of increasing stress—perceived stress, negative emotions, and depressive cognitions all increase. While this study does not purport or support the role of biofeedback directly leading to psychological changes, it does illustrate and validate the relationship between coherence and cognitive and emotional states.

By omitting the three studies reported separately, and combining the three Prinsloo articles into one study, the total number of studies evaluated in the following sections is 17. (See results below for detailed descriptions of full results, separated into various categories of interest, and refer to Table 2.1 at the end of the manuscript for comprehensive results of all studies.)

Populations Studied

Of the 17 studies reviewed, there was a variety of populations recruited. For the purposes of my evaluation and discussion, I have separated the types of populations into convenience samples (N=9; 53%) and deliberately selected samples (N=8; 47%). Convenience samples included high school students (Bradley et al., 2010; P. W. Kim, Kim, & Jung, 2012), college students (Astor, Adam, Jerčić, Schaaff, & Weinhardt, 2013; Escolano, Navarro-Gil, García-Campayo, & Minguez, 2014; Henriches, Keffer, Abrahamson, & Horst, 2011; Raaijmakers et al., 2013; Whited, Larkin, & Whited, 2014), and healthy adults (Cohen, Brinkman, & Neerincx, 2016; L. Sherlin, Muench, & Wyckoff, 2010). Deliberate samples included correctional officers (McCraty, Atkinson, Lipsenthal, & Arguelles, 2009), senior managers (Prinsloo et al., 2011;
Prinsloo, Derman, et al., 2013; Prinsloo, Rauch, et al., 2013), soldiers (Bouchard, Bernier, Boivin, Morin, & Robillard, 2012), professional male baseball players (L. H. Sherlin, Larson, & Sherlin, 2013), post-partum mothers (Kudo, Shinohara, & Kodama, 2014), police officers (J. P. Andersen & Gustafsberg, 2016), young male athletes (Dziembowska et al., 2016), and elite professional male soccer players (Rusciano, Corradini, & Stoianov, 2017). Individuals comprising samples in this latter category represent those whose optimal performance under acute stress is critical to execute their job successfully (Table 2.2).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study population</th>
<th>Category of population</th>
<th>Sample (Expr/Ctrl)</th>
<th>Mean age (StDev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCraty et al (2009)</td>
<td>Correctional officers</td>
<td>Deliberate</td>
<td>75 (43/32)</td>
<td><strong>Expr:</strong> 39.47 (7.70) <strong>Ctrl:</strong> 40.72 (8.12)</td>
</tr>
<tr>
<td>Bradley et al (2010)</td>
<td>High school students</td>
<td>Convenience</td>
<td>136 (77/59)</td>
<td><strong>Expr:</strong> 15.3 (0.44) <strong>Ctrl:</strong> 15.3 (0.44)</td>
</tr>
<tr>
<td>Sherlin et al (2010)</td>
<td>Healthy adults</td>
<td>Convenience</td>
<td>43</td>
<td>33.2 (8.77)</td>
</tr>
<tr>
<td>Henriques et al (2011)</td>
<td>College students</td>
<td>Convenience</td>
<td>35</td>
<td>Not reported</td>
</tr>
<tr>
<td>Prinsloo et al (2011, 2013a, 2013b)</td>
<td>Senior managers</td>
<td>Deliberate</td>
<td>18 (9/9)</td>
<td><strong>Expr:</strong> 19.2 (0.45)</td>
</tr>
<tr>
<td>Kim et al (2012)</td>
<td>High school students</td>
<td>Convenience</td>
<td>46 (21/25)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Astor et al (2013)</td>
<td>College students</td>
<td>Convenience</td>
<td><strong>Study 1:</strong> 36 (19/17); <strong>Study 2:</strong> 68 (44/24)</td>
<td><strong>Study 1:</strong> 23.39 (range: 8); <strong>Study 2:</strong> 22.06 (range: 9)</td>
</tr>
<tr>
<td>Sherlin et al (2013)</td>
<td>Professional male baseball players</td>
<td>Deliberate</td>
<td>5</td>
<td>19.2 (0.45)</td>
</tr>
<tr>
<td>Raaijmakers et al (2013)</td>
<td>College students</td>
<td>Convenience</td>
<td>28 (16/12)</td>
<td>22 (range: 8)</td>
</tr>
<tr>
<td>Escolano et al (2014)</td>
<td>College students</td>
<td>Convenience</td>
<td>19 (10/9)</td>
<td><strong>Expr:</strong> 25.8 (4.1) <strong>Ctrl:</strong> 24.3 (3.7)</td>
</tr>
<tr>
<td>Kudo et al (2014)</td>
<td>Post-partum mothers</td>
<td>Deliberate</td>
<td>55 (25/30)</td>
<td><strong>Expr:</strong> 30.5 (5.7) <strong>Ctrl:</strong> 33.4 (6.6)</td>
</tr>
<tr>
<td>Whited et al (2014)</td>
<td>Undergraduate and graduate students</td>
<td>Convenience</td>
<td>28 (15/13)</td>
<td><strong>Expr:</strong> 22.29 (3.6) <strong>Ctrl:</strong> 23.15 (4.06)</td>
</tr>
<tr>
<td>Dziembowska et al (2016)</td>
<td>Young male athletes</td>
<td>Deliberate</td>
<td>41 (20/21)</td>
<td>18.34 (1.36)</td>
</tr>
</tbody>
</table>
Table 2.2 Details on the study sample and population

In the following results, participant groups referred to as “experimental” include participants who received a biofeedback intervention, while groups referred to as “control” received either sham feedback, or no feedback at all. One study had an experimental group only (Sherlin et al., 2013), one study had a within-subjects design where each participant served as their own control (Cohen et al., 2016), and another study used an immediate and delayed intervention design (Henriques et al., 2011). In this design, each group served as a control group at one time point and an experimental group at the other time point, experiencing the same intervention at different times.

Types of Biofeedback Used

Across all studies included, 100% used the visual modality to convey biofeedback information, and 12% (N=2) additionally used the auditory modality to reinforce biofeedback (Bouchard et al., 2012; Sherlin et al., 2013). Both studies using visual and auditory biofeedback recruited specialized populations and delivered biofeedback in a way that abstractly represented physiological processes.

Since all studies, including these two, used the visual modality to represent the user’s physiology, I will further explore how this information was displayed. Overall, there were four total studies (24%) that described an abstract visual representation of underlying physiological state and changes in state (Astor et al., 2013; Bouchard et al., 2012; Escolano et al., 2014; Sherlin et al., 2013).
The most pervasive way of representing participants’ physiological state was by displaying direct physiological signals through the visual modality. This was done in 12 of the 17 studies (71%), and in each case, this signal was at least either HR, HRV, or HRV waveform (Bradley et al., 2010; Cohen et al., 2016; Dziembowska et al., 2016; Henriques et al., 2011; Kim et al., 2012; Kudo et al., 2014; McCraty et al., 2009; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Raaijmakers et al., 2013; Rusciano et al., 2017; Sherlin et al., 2010; Whited et al., 2014). Three of these 12 studies (25%; 18% of total 17 studies) presented the direct physiological signal, either as HR/HRV alone or in combination with other visual information. Of the 17 total studies, Cohen and colleagues were the only researchers to display direct indications of performance or error on the task (Cohen et al., 2016).

The remaining nine studies that presented physiology directly (75%; 53% of total 17 studies) displayed the HRV waveform and HRV biofeedback training geared towards achieving coherence (Bradley et al., 2010; Dziembowska et al., 2016; Henriques et al., 2011; Kudo et al., 2014; McCraty et al., 2009; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Rusciano et al., 2017; Sherlin et al., 2010; Whited et al., 2014). Whited and colleagues (2014) additionally displayed HR in BPM and the HRV power spectrum to participants, while Rusciano and colleagues (2017) included information regarding skin conductance level (SCL), electromyography signal, and skin temperature signal in addition to the HRV waveform and cognitive exercises. The remaining study provided no details on the visual display (Andersen et al., 2016; Table 2.3).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Modality: Visual</th>
<th>Category of visual display</th>
<th>Modality: Auditory</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCraty et al (2009)</td>
<td>HRV waveform (coherence)</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Bradley et al (2010)</td>
<td>HRV waveform (coherence)</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Authors (Year)</td>
<td>Methodology</td>
<td>Direct Physiological Signal(s)</td>
<td>Note(s)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Sherlin et al (2010)</td>
<td>HRV waveform (coherence)</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Henriques et al (2011)</td>
<td>HRV waveform (coherence)</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Prinsloo et al (2011, 2013a, 2013b)</td>
<td>HRV waveform (coherence)</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Bouchard et al (2012)</td>
<td>Red texture partially obscuring user’s field of view. Increased arousal represented by an increasingly obstructed view</td>
<td>Abstract representation(s)</td>
<td>Sound of a pumping heart increasing in frequency and loudness with increased arousal</td>
</tr>
<tr>
<td>Kim et al (2012)</td>
<td>ECG tachogram, anxiety profile</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Astor et al (2013)</td>
<td>Color-coded arousal meter; corresponding game elements changing color to reflect the state of arousal, located directly in attentional field</td>
<td>Abstract representation(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Sherlin et al (2013)</td>
<td>Dual Drive Xtreme: car presented in a race, moves at a speed proportional to user success; or Particle Editor: shapes or lines in movement along with music when parameters are met</td>
<td>Abstract representation(s)</td>
<td>Music</td>
</tr>
<tr>
<td>Raaijmakers et al (2013)</td>
<td>Across 6 tasks: 3 displayed SCL signal, 2 displayed HRV signal, 1 displayed a combination of HRV and SCL signals</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Escolano et al (2014)</td>
<td>A square with changing saturation colors; physiological values reflecting baseline appear gray, values above baseline appear red and increase in saturation as values increase, values below baseline appear blue and increase in saturation as values decrease</td>
<td>Abstract representation(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Kudo et al (2014)</td>
<td>HRV waveform (coherence)</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Whited et al (2014)</td>
<td>HR, HRV power spectrum, HRV waveform (coherence)</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Andersen et al (2016)</td>
<td>No details</td>
<td>No details</td>
<td>N/A</td>
</tr>
<tr>
<td>Cohen et al (2016)</td>
<td>Performance prediction bar graph, HR bar graph, error</td>
<td>Direct physiological signal(s)</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Physiological Equipment

Of the nine studies presenting the participants’ HRV waveform back during task completion, five of those (56%; 29% of total 17 studies) used Freeze-Framer/emWave to display physiological information (Bradley et al., 2010; Dziembowska et al., 2016; Henriques et al., 2011; McCraty, Atkinson, Lipsenthal, et al., 2009; Whited et al., 2014). Three of these studies (33%; 18% of total 17 studies) used StressEraser to display information back to the participant (Kudo et al., 2014; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2010). Most of these studies used the devices listed to collect, process, and display information in real time, but Prinsloo and colleagues (2011, 2013a, 2013b) collected and processed data using BioPac equipment and fed that data back using StressEraser, while Whited and colleagues (2014) collected and processed data using Polar RS800CS equipment and fed that data back using emWave. The remaining study that presented HRV waveform as the visual indicator of physiology used the Nexus 10 Mark II system to collect, process, and present physiological data (Rusciano et al., 2017).

Five of the total 17 studies (29%) used chest-worn HR belts to collect HR data (Andersen et al., 2016; Astor et al., 2013; Bouchard et al., 2012; Cohen et al., 2016; Whited et al., 2014).

Table 2.3 Details on the presentation structure of visual and auditory forms of biofeedback (HRV: heart rate variability; ECG: electrocardiography; SCL: skin conductance level; EMG: electromyography)

<table>
<thead>
<tr>
<th>Study</th>
<th>Presentation Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dziembowska et al (2016)</td>
<td>HRV waveform (coherence)</td>
</tr>
<tr>
<td></td>
<td>Direct physiological signal(s)</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Rusciano et al (2017)</td>
<td>Sessions 1-3: HRV waveform (coherence); Sessions 4-9: HRV waveform (coherence), SCL level, EMG signal, skin temperature signal; Sessions 10-15: HRV waveform (coherence)</td>
</tr>
<tr>
<td></td>
<td>Direct physiological signal(s)</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>
Zephyr (Andersen et al., 2016; Cohen et al., 2016) and Polar (Bouchard et al., 2012; Whited et al., 2014) were equally popular among these five studies, with one remaining unspecified (Astor et al., 2013; Table 2.4).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Physio measures</th>
<th>Physio equipment</th>
<th>Physio results</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCratty et al (2009)</td>
<td>Cortisol, DHEA, S-IgA cholesterol, triglycerides, fasting glucose levels, HR, HRV, BP</td>
<td>Freeze-Framer/emWave</td>
<td>Pre-post Expr: ↓ total cholesterol, LDL, total/HDL ratio, glucose levels, DHEA, mean arterial pressure, systolic BP, diastolic BP, mean HR, ↑ LF/HF ratio of HRV; Pre-post Ctrl: ↓ total cholesterol, HDL, glucose levels</td>
</tr>
<tr>
<td>Sherlin et al (2010)</td>
<td>QEEG</td>
<td>StressEraser</td>
<td>Pre-post Expr: ↑ overall alpha; Pre-post Ctrl: ↓ overall alpha</td>
</tr>
<tr>
<td>Henriques et al (2011)</td>
<td>HRV</td>
<td>emWave</td>
<td>Pre-post Expr: no significant differences</td>
</tr>
<tr>
<td>Kim et al (2012)</td>
<td>HRV</td>
<td>QECG-3</td>
<td>Pre-post Expr: ↑ R-R intervals</td>
</tr>
<tr>
<td>Astor et al (2013)</td>
<td>HR (arousal level)</td>
<td>Chest-worn HR belt</td>
<td>Study 1 Expr vs. Ctrl: not reported; Study 2 Expr vs. Ctrl: ↓ arousal level</td>
</tr>
<tr>
<td>Raaijmakers et al (2013)</td>
<td>HRV, SCL, EEG alpha asymmetry</td>
<td>ActiveTwo</td>
<td>Pre-post Expr vs. Ctrl: No significant differences</td>
</tr>
<tr>
<td>Kudo et al (2014)</td>
<td>HRV</td>
<td>StressEraser</td>
<td>Pre-post Expr vs. Ctrl: ↓ HR, ↑ SDNN, HF power</td>
</tr>
<tr>
<td>Study</td>
<td>Physiological Equipment</td>
<td>Measures</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Andersen et al (2016)</td>
<td>HR, BP</td>
<td>Zephyr BioHarness, Omniscense ambulatory blood pressure wrist cuffs</td>
<td>Post Expr vs. Ctrl: ↓ HR max between baseline and Scenario 1, ↓ change in HR from baseline to Scenario 1</td>
</tr>
<tr>
<td>Cohen et al (2016)</td>
<td>HR, HRV</td>
<td>Zephyr HxM</td>
<td>Pre-post Expr: ↑ HRV spectral power, LF HRV, IC, theta and alpha spectral power, alpha asymmetry, ↓ HF HRV</td>
</tr>
<tr>
<td>Dziembowska et al (2016)</td>
<td>EEG alpha asymmetry, HRV</td>
<td>emWave, Electro-Cap International</td>
<td></td>
</tr>
<tr>
<td>Rusciano et al (2017)</td>
<td>SCL, RR, BVP, HR, HRV, surface EMG, skin temperature</td>
<td>Nexus 10 Mark II</td>
<td>Post Expr vs. Ctrl: ↓ SCL, HR, RR; Pre-post Expr: ↓ RR ↑ LF</td>
</tr>
</tbody>
</table>

Table 2.4 Details regarding physiological equipment, measures, and results (DHEA: dehydroepiandrosterone; S-IgA: secretory immunoglobulin; HR: heart rate; HRV: heart rate variability; BP: blood pressure; LDL: low-density lipoprotein cholesterol; HDL: high-density lipoprotein cholesterol; LF: low frequency power; HF: high frequency power; SDNN: standard deviation from normal-to-normal; IC: coherence index; QEEG: quantitative electroencephalography; EEG: electroencephalography; RR: respiratory rate; RF: respiratory frequency; TF: total frequency power; RMSSD: root mean square of the successive differences; QECG-3: Quantitative Electrocardiography-Three Limb Lead; SCL: skin conductance level; UA: upper alpha bandwidth; LA2: lower alpha 2 bandwidth; pNN50: percentage of normal-to-normal peaks with a difference greater than 50 milliseconds; BVP: blood volume pulse; EMG: electromyography). In the Results column, “Pre-post Expr” refers to a main effect of time, where the experimental group experienced the associated significant changes in physiological measures (in the direction indicated by the arrow) after receiving a biofeedback intervention compared to beforehand. “Pre-post Ctrl” indicates that significant differences were observed in the control group at the second time point compared to the first time point, despite the lack of a biofeedback intervention in between. “Post Expr vs. Ctrl” indicates a main effect of group was reported, such that the experimental group saw significant changes compared to the control group at the post-intervention time point. “Pre-post Expr vs. Ctrl” refers to the presence of an interaction effect, where the experimental group experienced significantly different results after the intervention compared to the control group. “Interv Expr vs. Ctrl,” appearing only once (Prinsloo et al., 2011; Prinsloo et al., 2013b), refers to a main effect of group where the physiological measures acquired during the intervention were significantly different for the experimental group compared to the control group. One study (Bouchard et al., 2012) reported a main effect of group such that the experimental group had significantly different values compared to the control group at a mid-point in the design (referred to as “Mid Expr vs. Ctrl”). “Pre-post,” also appearing only once (Sherlin et al., 2013), refers to the lack of a control group.

Physiological Measures and Results

Across the 17 studies reviewed, there was a wide variety of physiological measures collected and analyzed by researchers (Table 2.5). The most common were HR/HRV, measured...
in 14 of the 17 total studies (82%), and electroencephalography (EEG), measured in 6 of the 17 (35%).

Of the 14 HR/HRV studies, 11 (79%) reported significant differences in the experimental group in at least one cardiovascular measure (Andersen et al., 2016; Astor et al., 2013; Bouchard et al., 2012; Bradley et al., 2010; Dziembowska et al., 2016; Kim et al., 2012; Kudo et al., 2014; McCraty et al., 2009; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Rusciano et al., 2017; Whited et al., 2014). Four of these 11 studies reported a main effect of time, such that the experimental group saw significant changes at the post-intervention time point compared to the pre-intervention time point (Dziembowska et al., 2016; Kim et al., 2012; McCraty, Atkinson, Lipsenthal, et al., 2009; Rusciano et al., 2017). Four studies also reported a main effect of group, such that the experimental group saw significantly different changes at the post-intervention time point (Andersen & Gustafsberg, 2016; Rusciano et al., 2017) or during the intervention (Astor et al., 2013; Prinsloo et al., 2011; Prinsloo et al., 2013b). The remaining four studies reported interaction effects, such that HR/HRV values saw significant changes among the experimental group compared to the control group at the post-intervention time point compared to a baseline or pre-intervention time point (Bouchard et al., 2012; Bradley et al., 2010; Kudo et al., 2014; Whited et al., 2014). Two studies (14%) reported no significant differences in HR/HRV (Henriques et al., 2011; Raaijmakers et al., 2013), and one study (7%) did not report the data (Cohen et al., 2016).

Other HR/HRV indices reported with significant changes include: arousal level (Astor et al., 2013); R-R intervals (Kim et al., 2012); low frequency/high frequency ratio (McCraty et al., 2009); root mean square of the successive differences (RMSSD; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b); percentage of normal-to-normal peaks with a difference
greater than 50 milliseconds (pNN50; Whited et al., 2014); coherence index (Bradley et al.,
2010; Dziembowska et al., 2016); standard deviation from normal-to-normal (SDNN; Bradley et
al., 2010; Kudo et al., 2014; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b);
and low frequency power (Bradley et al., 2010; Dziembowska et al., 2016; Rusciano et al.,
2017). See Table 2.4 for a detailed account of physiological measures and results.

Of the six studies measuring EEG, whether using quantitative EEG (Sherlin et al., 2013;
Sherlin et al., 2010), traditional EEG (Escolano et al., 2014; Prinsloo et al., 2011; Prinsloo et al.,
2013a; Prinsloo et al., 2013b), or alpha asymmetry (Dziembowska et al., 2016; Raaijmakers et
al., 2013), five saw significant differences (83%) within at least one component in the
experimental group compared to the control group. Raaijmakers and colleagues were the only
group reporting no significant differences (Raaijmakers et al., 2013). Each of the five studies
reporting significant differences between experimental and control groups reported changes in
distinct bandwidths and neural areas.

Of the measures collected and evaluated in only one study, many of those showed
statistically significant improvements when comparing levels before the intervention to
afterwards, in the experimental group alone. These measures included dehydroepiandrosterone,
secretory immunoglobulin A, cholesterol, triglycerides, fasting glucose levels (McCraty,
Atkinson, Lipsenthal, et al., 2009), respiratory frequency (Prinsloo et al., 2011; Prinsloo et al.,
2013a; Prinsloo et al., 2013b), and surface electromyography, skin temperature, and blood
volume pulse (BVP; Rusciano et al., 2017).

Additionally, cortisol (Bouchard et al., 2012; McCraty, Atkinson, Lipsenthal, et al.,
2009), respiratory rate (RR; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b;
Rusciano et al., 2017), and SCL (Raaijmakers et al., 2013; Rusciano et al., 2017) were evaluated
in two studies each (12%), while blood pressure (BP) was evaluated in three studies (18%; Andersen et al., 2016; McCraty et al., 2009; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b). Cortisol levels were significantly reduced in the experimental group compared to the control group in one of the two studies in which it was evaluated (Bouchard et al., 2012). RR was also significantly reduced in the experimental group compared to the control group, in both studies (Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Rusciano et al., 2017). Analysis of SCL showed a significant reduction in the experimental group when comparing levels before and after the intervention, reported in one of the two studies evaluating SCL (Rusciano et al., 2017). BP decreased significantly within the experimental group compared to the control group in one of the three studies (McCraty, Atkinson, Lipsenthal, et al., 2009).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physiological Measures</strong></td>
<td></td>
</tr>
<tr>
<td>Heart rate</td>
<td>Average heartbeats per minute over a time window</td>
</tr>
<tr>
<td>R-R intervals</td>
<td>Variability among R-R intervals from ECG data provide heart rate</td>
</tr>
<tr>
<td>Low frequency power</td>
<td>HRV power spectrum ranging from 0.04 to 0.15 Hz</td>
</tr>
<tr>
<td>High frequency power</td>
<td>HRV power spectrum ranging from 0.15 to 0.4 Hz</td>
</tr>
<tr>
<td>Low frequency/High frequency ratio</td>
<td>Ratio of low frequency power to high frequency power</td>
</tr>
<tr>
<td>Total frequency power</td>
<td>HRV power spectrum ranging from 0.005 to 0.4 Hz</td>
</tr>
<tr>
<td>Root mean square of the successive differences</td>
<td>HRV time-domain measure calculating the variation among successive R-R intervals</td>
</tr>
<tr>
<td>Standard deviation from normal-to-normal</td>
<td>HRV time-domain measure calculating the variation of R-R interval values in reference to a mean</td>
</tr>
<tr>
<td>pNN50</td>
<td>HRV time-domain measure calculating the percentage of successive R-R intervals that differ by 50 ms or more</td>
</tr>
<tr>
<td>Coherence index</td>
<td>Primary marker of psychophysiological coherence state, derived from the degree of symmetry of the HRV waveform; a higher value indicates higher symmetry and a more positive state</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Delta bandwidth</td>
<td>EEG power spectrum ranging from 1 to 3.5 Hz</td>
</tr>
<tr>
<td>Theta bandwidth</td>
<td>EEG power spectrum ranging from 4 to 7.5 Hz</td>
</tr>
<tr>
<td>Alpha bandwidth</td>
<td>EEG power spectrum ranging from 8 to 12 Hz</td>
</tr>
<tr>
<td>Beta bandwidth</td>
<td>EEG power spectrum ranging from 13 to 32 Hz</td>
</tr>
<tr>
<td>Gamma bandwidth</td>
<td>EEG power spectrum ranging from 32 to 45 Hz</td>
</tr>
<tr>
<td>Theta/Beta ratio</td>
<td>Ratio of theta spectral power to beta spectral power</td>
</tr>
<tr>
<td>Skin conductance level</td>
<td>Measure of the degree of electrical conduction across the skin, determined by activity of eccrine sweat glands</td>
</tr>
<tr>
<td>Respiratory rate</td>
<td>Average breaths per minute over a time window</td>
</tr>
<tr>
<td>Cortisol</td>
<td>Stress-related hormone, measurable by saliva or blood sample collection</td>
</tr>
<tr>
<td>Low-density lipoprotein (LDL) cholesterol</td>
<td>Transporter of fat molecules; contributes to fatty build-up in arteries when elevated</td>
</tr>
<tr>
<td>High-density lipoprotein (HDL) cholesterol</td>
<td>Transporter of LDL cholesterol from arteries to the liver, where it is broken down</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>Most common type of fat in the body; when measured in blood, high levels are an indicator of poor health</td>
</tr>
<tr>
<td>Total cholesterol</td>
<td>Measure of overall cholesterol levels, derived by summing LDL, HDL and triglyceride levels</td>
</tr>
<tr>
<td>Blood glucose levels</td>
<td>Overall level of sugar in the blood; high levels are an indicator of poor health</td>
</tr>
<tr>
<td>Dehydroepiandrosterone</td>
<td>Stress-related hormone</td>
</tr>
<tr>
<td>Secretory immunoglobulin A</td>
<td>Mucosal class of antibody with protective immunological qualities</td>
</tr>
<tr>
<td>Mean arterial pressure</td>
<td>The average blood pressure within an individual during one cardiac cycle</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>Total pressure in blood vessels while the heart beats</td>
</tr>
<tr>
<td>Test Name</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Diastolic blood pressure</td>
<td>Total pressure in blood vessels between heartbeats</td>
</tr>
<tr>
<td><strong>Psychological measures</strong></td>
<td></td>
</tr>
<tr>
<td>Brief Symptom Inventory</td>
<td>53-item questionnaire measuring psychological symptom patterns with 9 symptom scales: Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism; and 3 global indices: Global Severity Index, Positive Symptom Distress Index, and Positive Symptom Total</td>
</tr>
<tr>
<td>State Trait Anxiety Inventory for Adults</td>
<td>20-item questionnaire evaluating the stable trait of stress; optional 20-item questionnaire evaluating the momentary state of stress</td>
</tr>
<tr>
<td>Personal Wellness Profile</td>
<td>75-item questionnaire assessing health information, physical activity, eating practices, alcohol, drugs, and smoking, stress and coping, social health, safety, medical care, and health view</td>
</tr>
<tr>
<td>Jenkins Activity Survey</td>
<td>Measure of Type A and coronary-prone behavior with three subscales: Speed and Impatience, Job Involvement, and Hard-Driving and Competitive</td>
</tr>
<tr>
<td>Personal and Organizational Quality Assessment</td>
<td>Measure of psychological and workplace contributors to organizational climate with 2 main areas: Personal Scales and Organizational Scales</td>
</tr>
<tr>
<td>Test Anxiety Inventory</td>
<td>16-item questionnaire measuring test anxiety globally, with 2 additional measurements of Worry and Emotionality</td>
</tr>
<tr>
<td>Perceived Stress Scale</td>
<td>10-item questionnaire assessing the global measure of self-perceived stress</td>
</tr>
<tr>
<td>Mood and Anxiety Symptom Questionnaire</td>
<td>90-item questionnaire measuring mood and anxiety symptoms with 6 subscales: General Distress Mixed, General Distress Anxious, General Distress Depressed, Anxious Arousal, Loss of Interest, and High Positive Affect</td>
</tr>
<tr>
<td>Scales of Psychological Well-Being</td>
<td>54-item questionnaire measuring aspects of psychological functioning with 6 subscales: Self-Acceptance, Positive Relations with Others, Autonomy, Environmental Mastery, Purpose in Life, and Personal Growth</td>
</tr>
<tr>
<td>Smith Relaxation States Inventory 3</td>
<td>38-item questionnaire assessing relaxation states with 19 total relaxation states falling into 4 categories: Mindfulness, Energized Positive Feelings, Basic Relaxation, and Transcendence</td>
</tr>
<tr>
<td>Perceived self-efficacy scale*</td>
<td>4-item questionnaire measuring confidence in oneself to control stress at different intensities</td>
</tr>
<tr>
<td>Simulator Sickness Questionnaire</td>
<td>16-item questionnaire measuring side effects induced by immersive simulation</td>
</tr>
<tr>
<td>Appreciation and usefulness of program*</td>
<td>13-item questionnaire evaluating perceptions of the intervention</td>
</tr>
</tbody>
</table>
The majority of studies reviewed, 14 of 17 (82%), included psychological assessments via self-report instruments. While there was some consistency in the types of measures collected, there was little overlap among the instruments used. The most frequently evaluated psychological measures included clinical symptoms, sleep, and stress/anxiety (Table 2.6, Table 2.1).

Clinical symptoms (McCraty, Atkinson, Lipsenthal, et al., 2009; Whited et al., 2014) and sleep (Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2013)
were each assessed across two studies (12% of those evaluating psychological measures). Both
studies evaluating clinical symptoms used the Brief Symptom Inventory (Derogatis et al., 2012),
and only one of those saw significant improvement from pre- to post-intervention assessment in
the experimental group (McCraty, Atkinson, Lipsenthal, et al., 2009). Sleep was evaluated
differently in each of the studies where it was reported, but in both cases there was improvement
(Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2013). This
improvement was demonstrated by an interaction effect in one study, whereby the experimental
group reported less sleepiness after the intervention versus before, compared to the control group
(Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo, et al., 2013b), and as a main effect of time
for the experimental group in one other study (Sherlin et al., 2013).

The most consistently evaluated measure was stress/anxiety, assessed in 8 of the 14
studies (57%) measuring psychological constructs (Andersen et al., 2016; Bradley et al., 2010;
Dziembowska et al., 2016; Henriques et al., 2011; Prinsloo et al., 2011; Prinsloo et al., 2013a;
Prinsloo et al., 2013b; Sherlin et al., 2013; Sherlin et al., 2010; Whited et al., 2014). Four of
these eight studies used the State Trait Anxiety Inventory for Adults (Spielberger et al., 1983) to
assess the construct of stress (Dziembowska et al., 2016; Henriques et al., 2011; Prinsloo et al.,
2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2010). The remaining four
studies evaluating stress/anxiety used unique instruments. Of the eight studies evaluating
stress/anxiety, five of these (62.5%) saw a significant decrease in stress and/or anxiety from pre-
to post-intervention in the experimental group (Bradley et al., 2010; Dziembowska et al., 2016;
Henriques et al., 2011; Sherlin et al., 2013; Sherlin et al., 2010).
Aside from the instruments mentioned, there were no additional standardized assessments appearing in more than one study. Overall, there were 23 unique scales or inventories used across the 14 studies that evaluated psychological constructs (Table 2.6).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Psych measures</th>
<th>Psych instruments</th>
<th>Psych results</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCraty et al (2009)</td>
<td>Health risk, Type A and coronary-prone behavior, psychological symptoms, psychological and workplace elements</td>
<td>Personal Wellness Profile, Jenkins Activity Survey, Brief Symptom Inventory, Personal and Organizational Quality Assessment</td>
<td><strong>Pre-post Expr:</strong> ↓ anger, fatigue, hostility, interpersonal sensitivity, paranoid ideation, psychological distress, Type A behavior, speed and impatience, ↑ productivity, motivation, goal clarity, perceived manager support, gratitude, positive outlook; <strong>Pre-post Ctrl:</strong> ↓ work attitude, confidence in organization, ↑ depression</td>
</tr>
<tr>
<td>Bradley et al (2010)</td>
<td>Test anxiety (global, worry, emotionality)</td>
<td>Test Anxiety Inventory (8 of 16 items)</td>
<td><strong>Pre-post Expr vs. Ctrl:</strong> ↓ overall test anxiety (global, worry, emotionality)</td>
</tr>
<tr>
<td>Sherlin et al (2010)</td>
<td>State anxiety</td>
<td>State-Trait Anxiety Interview- State Form</td>
<td><strong>Pre-post Expr vs. Ctrl:</strong> ↓ state anxiety</td>
</tr>
<tr>
<td>Henriques et al (2011)</td>
<td>Mood/anxiety symptoms, state and trait anxiety, subjective well-being</td>
<td>Mood and Anxiety Symptom Questionnaire, State Trait Anxiety Inventory, Scales of Psychological Well-Being</td>
<td><strong>Pre-post Expr:</strong> ↓ general distress: mixed, general distress: anxious, anxious arousal</td>
</tr>
<tr>
<td>Prinsloo et al (2011, 2013a, 2013b)</td>
<td>State and trait anxiety, relaxation, sleepiness</td>
<td>State-Trait Anxiety Inventory Forms Y-1 and Y-2, Smith Relaxation States Inventory 3, Visual Analog Scale</td>
<td><strong>Pre-post Expr vs Ctrl:</strong> ↑ relaxation, ↓ sleepiness</td>
</tr>
<tr>
<td>Astor et al (2013)</td>
<td>Cognitive reappraisal and suppression</td>
<td>Emotion Regulation Questionnaire</td>
<td>Not reported</td>
</tr>
<tr>
<td>Sherlin et al (2013)</td>
<td>Focus index, stress index, speed index</td>
<td>NeuroPerformance Profile</td>
<td><strong>Pre-post:</strong> ↑ sleep, focus, feeling of being relaxed, all indices of NeuroPerformance Profile</td>
</tr>
<tr>
<td>Raaijmakers et al (2013)</td>
<td>Affect</td>
<td>Positive and Negative Affect Scale</td>
<td><strong>Pre-post Expr vs. Ctrl:</strong> No significant differences</td>
</tr>
<tr>
<td>Study</td>
<td>Psychological Constructs</td>
<td>Psychological Instruments/Measures</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Whited et al (2014)</td>
<td>Distress, stress, state affect</td>
<td>Brief Symptom Inventory, Perceived Stress Scale, Multiple Adjective Affect Checklist-Revised</td>
<td>Pre-post Expr vs. Ctrl: No significant differences</td>
</tr>
<tr>
<td>Andersen et al (2016)</td>
<td>Perceived stress, skills, ability to manage scenarios, self-confidence in decision-making, situational awareness</td>
<td>Stress, Skills, Ability to manage scenarios, Situational awareness</td>
<td>Pre Ctrl vs. Expr: ↑ confidence</td>
</tr>
<tr>
<td>Cohen et al (2016)</td>
<td>Appraisal, task demand, usability</td>
<td>Likert-type scale, Level-of-Information Processing Scale, System Usability Scale</td>
<td>HR vs. noHR: ↓ perceived usability of feedback</td>
</tr>
<tr>
<td>Dziembowska et al</td>
<td>State anxiety, self esteem</td>
<td>State-Trait Anxiety Inventory A-State Scale, Rosenberg Self Esteem Scale</td>
<td>Pre-post Expr: ↓ state anxiety</td>
</tr>
</tbody>
</table>

**Table 2.6** Details regarding psychological instruments, measures, and results for the studies assessing psychological constructs. In the Results column, “Pre-post Expr” refers to a main effect of time, where the experimental group experienced the associated significant changes on self-reported measures (in the direction indicated by the arrow) after receiving a biofeedback intervention compared to beforehand. “Pre-post Ctrl” indicates that significant differences were observed in the control group at the second time point compared to the first time point, despite the lack of a biofeedback intervention in between. “Post Expr vs. Ctrl” indicates a main effect of group was reported, such that the experimental group saw significant changes compared to the control group at the post-intervention time point. One study reported a main effect of group reflecting an significantly higher value self-reported measure in the control group at the pre-intervention time point (noted as “Pre Ctrl vs. Expr”; Andersen et al., 2016). “Pre-post Expr vs. Ctrl” refers to the presence of an interaction effect, where the experimental group experienced significantly different results after the intervention versus beforehand, compared to the control group. “HR vs. noHR,” appearing only once (Cohen et al., 2016), refers to a main effect of group where the self-reported responses recorded while heart rate feedback was available were significantly different for that same group of participants in a different condition while heart rate feedback was not available (within-subjects experimental design). “Pre-post,” also appearing only once (Sherlin et al., 2013), refers to the lack of a control group.
Performance Instruments, Measures, and Results

Twelve of the 17 total studies (71%) evaluated performance (Andersen et al., 2016; Astor et al., 2013; Bouchard et al., 2012; Bradley et al., 2010; Cohen et al., 2016; Escolano et al., 2014; Kim et al., 2012; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Raaijmakers et al., 2013; Rusciano et al., 2017; Sherlin et al., 2013; Sherlin et al., 2010). Across these 12 studies, there were only two performance indices and one assessment that was repeated more than once.

While 8 of the 12 studies evaluating performance (67%) used at least reaction time (RT) and/or errors made to approximate task performance (Astor et al., 2013; Cohen et al., 2016; Escolano et al., 2014; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Raaijmakers et al., 2013; Rusciano et al., 2017; Sherlin et al., 2013; Sherlin et al., 2010), five total studies (42%) evaluated performance in a way that was more specialized to the populations recruited (Andersen et al., 2016; Bouchard et al., 2012; Bradley et al., 2010; Kim et al., 2012; Rusciano et al., 2017). Of those evaluating RT and errors made, four (50%) used a version of the Stroop task (Escolano et al., 2014; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Rusciano et al., 2017; Sherlin et al., 2010). The other four studies evaluating performance via RT and errors used unique tasks to measure those metrics (Astor et al., 2013; Cohen et al., 2016; Raaijmakers et al., 2013; Sherlin et al., 2013). Six out of these eight tasks (75%) saw improvements in some component as a result of the intervention (Astor et al., 2013; Cohen et al., 2016; Escolano et al., 2014; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Rusciano et al., 2017; Sherlin et al., 2013).

Four of the 12 studies assessing performance (33%) included performance measures or tasks related to the deliberate population recruited (Andersen et al., 2016; Bouchard et al., 2012;
Rusciano et al., 2017; Sherlin et al., 2013). Notably, these four studies comprise half of the total studies recruiting specialized or deliberate samples. All four of these studies reported improvements in some aspect of performance as a result of the intervention. In contrast, three of the 12 studies (25%) saw no improvements (Bradley et al., 2010; Raaijmakers et al., 2013; Sherlin et al., 2010), all of which recruited participants using convenience sampling methods. Overall, nine of the 12 studies evaluating performance (75%) saw some improvements (Andersen et al., 2016; Astor et al., 2013; Bouchard et al., 2012; Cohen et al., 2016; Escolano et al., 2014; Kim et al., 2012; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Rusciano et al., 2017; Sherlin et al., 2013; Table 2.7).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Performance measures</th>
<th>Performance instruments</th>
<th>Performance results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bradley et al (2010)</td>
<td>Standardized testing scores</td>
<td>California High School Exit Exam, California Standards Test</td>
<td>Pre-post Expr vs. Ctrl: No significant differences</td>
</tr>
<tr>
<td>Prinsloo et al (2011, 2013a, 2013b)</td>
<td>RT, mistakes made</td>
<td>Stroop task</td>
<td>Pre-post Expr vs. Ctrl: ↑ reaction time, consistency in responses, ↓ mistakes made</td>
</tr>
<tr>
<td>Raaijmakers et al (2013)</td>
<td>RT, error percentages</td>
<td>N-back task, Mental Rotation Task</td>
<td>Pre-post Expr: ↓ RT, ↑ accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pre-post Ctrl: ↓ RT, ↑ accuracy</td>
</tr>
</tbody>
</table>
Table 2.7 Details regarding performance instruments, measures, and results for the studies assessing performance (RT: reaction time). In the Results column, “Pre-post Expr” refers to a main effect of time, where the experimental group experienced the associated significant changes in some aspect of performance (in the direction indicated by the arrow) after receiving a biofeedback intervention compared to beforehand. “Pre-post Ctrl” indicates that significant differences were observed in the control group at the second time point compared to the first time point, despite the lack of a biofeedback intervention in between. “Post Expr vs. Ctrl” indicates a main effect of group was reported, such that the experimental group saw significant changes compared to the control group at the post-intervention time point. “Pre-post Expr vs. Ctrl” refers to the presence of an interaction effect, where the experimental group experienced significantly different results after the intervention versus beforehand, compared to the control group. One study used a within-subjects design with multiple experimental conditions and multiple control conditions per participant (Cohen et al., 2016). Results are presented as “Expr vs. Ctrl” to indicate all experimental conditions contrasted with all control conditions. “Pre-post,” also appearing only once (Sherlin et al., 2013), refers to the lack of a control group.

Features of the Intervention

Session duration had little variety across studies. Of the 17 studies, 13 total (76%) consisted of session durations that lasted for less than one hour (Astor et al., 2013; Bouchard et al., 2012; Dziembowska et al., 2016; Escolano et al., 2014; Henriques et al., 2011; Kim et al., 2012; McCraty et al., 2009; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b;
Raaijmakers et al., 2013; Rusciano et al., 2017; Sherlin et al., 2013; Sherlin et al., 2010; Whited et al., 2014), with the longest duration of those being 35 minutes (Sherlin et al., 2013). Only one of the studies was longer than one hour (6%), ranging from 90 to 120 minutes (Cohen et al., 2016). Additionally, one study (6%) reported the session duration as variable, depending on how long it took the participant to achieve a certain objective (Kudo et al., 2014). Two of the 17 studies (12%), however, provided no detail regarding session duration (Andersen et al., 2016; Bradley et al., 2010).

The frequency of sessions over the course of the study was more variable across studies. The experimental session for five of the 17 studies (29%) occurred only once (Astor et al., 2013; Cohen et al., 2016; Escolano et al., 2014; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2010). Eleven of the 17 studies (65%) reported multiple sessions (Bouchard et al., 2012; Bradley et al., 2010; Dziembowska et al., 2016; Escolano et al., 2014; Henriques et al., 2011; Kim et al., 2012; Kudo et al., 2014; McCraty et al., 2009; Raaijmakers et al., 2013; Rusciano et al., 2017; Sherlin et al., 2013; Whited et al., 2014). Of those studies with repeated sessions, five of the eleven (45%) had between three and ten total sessions (Bouchard et al., 2012; Dziembowska et al., 2016; Kim et al., 2012; Raaijmakers et al., 2013; Whited et al., 2014), while six of the eleven (55%) held 15 or more sessions (Bradley et al., 2010; Henriques et al., 2011; Kudo et al., 2014; McCraty et al., 2009; Rusciano et al., 2017; Sherlin et al., 2013), with the greatest number of sessions approaching 40 (Bradley et al., 2010). One study (6%) had no details regarding the frequency of experimental sessions (Andersen et al., 2016).

Despite the relative consistency among session duration and frequency of intervention across studies, there was a much greater variety in terms of study duration. Study duration ranged from one-day sessions to five months. Seven of the 17 total studies (41%) ran for a period
of days (Andersen et al., 2016; Astor et al., 2013; Bouchard et al., 2012; Cohen et al., 2016; Escolano et al., 2014; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2010). Three studies (18%) spanned for weeks (Dziembowska et al., 2016; Kim et al., 2012; Raaijmakers et al., 2013). The remaining seven studies (41%) ran over the course of months (Bradley et al., 2010; Henriques et al., 2011; Kudo et al., 2014; McCraty et al., 2009; Rusciano et al., 2017; Sherlin et al., 2013; Whited et al., 2014; Table 2.8).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Session</th>
<th>Frequency</th>
<th>Duration of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCraty et al (2009)</td>
<td>15-min Heart Lock-Ins,</td>
<td>Daily</td>
<td>3 months</td>
</tr>
<tr>
<td></td>
<td>3 FreezeFrames per week; 30-sec</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FreezeFrames/hour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bradley et al (2010)</td>
<td>TestEdge</td>
<td>2 times per week</td>
<td>5 months</td>
</tr>
<tr>
<td>Sherlin et al (2010)</td>
<td>15 minutes</td>
<td>Once</td>
<td>1 day</td>
</tr>
<tr>
<td>Henriques et al (2011)</td>
<td>15 minutes</td>
<td>5 times per week</td>
<td>1 month</td>
</tr>
<tr>
<td>Prinsloo et al (2011,</td>
<td>10 minutes</td>
<td>Once</td>
<td>1 day</td>
</tr>
<tr>
<td>2013a, 2013b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bouchard et al (2012)</td>
<td>30 minutes</td>
<td>Daily</td>
<td>3 days</td>
</tr>
<tr>
<td>Kim et al (2012)</td>
<td>15 minutes</td>
<td>3 total sessions</td>
<td>3 weeks</td>
</tr>
<tr>
<td>Astor et al (2013)</td>
<td>25 minutes</td>
<td>Once</td>
<td>1 day</td>
</tr>
<tr>
<td>Sherlin et al (2013)</td>
<td>20-35 minutes</td>
<td>Average of 3 sessions per week (15 total per participant)</td>
<td>1 month</td>
</tr>
<tr>
<td>Raaijmakers et al (2013)</td>
<td>30 minutes</td>
<td>7 sessions</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Escolano et al (2014)</td>
<td>25 minutes</td>
<td>1 session</td>
<td>3 days</td>
</tr>
<tr>
<td>Kudo et al (2014)</td>
<td>To a certain score (untimed)</td>
<td>At least 1 session per day</td>
<td>1 month</td>
</tr>
<tr>
<td>Whited et al (2014)</td>
<td>32 minutes</td>
<td>4 to 8 sessions</td>
<td>1-3 months</td>
</tr>
<tr>
<td>Andersen &amp; Gustafberg (2016)</td>
<td>iPREP</td>
<td>No details</td>
<td>6 days</td>
</tr>
<tr>
<td>Cohen et al (2016)</td>
<td>90-120 minutes</td>
<td>Once</td>
<td>1 day</td>
</tr>
<tr>
<td>Dziembowska et al (2016)</td>
<td>20 minutes</td>
<td>10 sessions</td>
<td>3 weeks</td>
</tr>
<tr>
<td>Rusciano et al (2017)</td>
<td>15-30 minutes</td>
<td>15 sessions</td>
<td>4 months</td>
</tr>
</tbody>
</table>

Table 2.8 Details about the intervention frequency and study duration (iPREP: International Performance Resilience and Efficiency Program)
Discussion

The results of this systematic review support a variety of physiological, psychological, and performance improvements when biofeedback interventions are administered for stress management. One of the 17 total studies evaluated reported a lack of statistically significant change across all measures (Raaijmakers et al., 2013). The only significant findings in this study were two main effects of time, indicating an improvement in performance in the experimental group from pre- to post-intervention, as well as an improvement in performance for the control group from pre- to post-intervention. All remaining studies reported at least one meaningful significant difference in the experimental group following the intervention among either physiological, psychological, or performance assessments. In many cases, significant observations were made in more than one domain, and in some cases, in all three (Andersen et al., 2016; Bouchard et al., 2012; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2013).

Who Benefits the Most from Biofeedback Interventions?

When we take a closer look and consider the results in regards to the populations enrolled, we can make a more nuanced conclusion about specifically for whom a biofeedback intervention for stress management might benefit most. All eight studies that deliberately recruited specialized populations observed benefits across at least one measure within every domain analyzed. For example, all eight studies collected and analyzed physiological measures, and all eight reported improvement in the key components collected (Andersen et al., 2016; Bouchard et al., 2012; Dziembowska et al., 2016; Kudo et al., 2014; McCraty et al., 2009; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Rusciano et al., 2017; Sherlin et al., 2013).
et al., 2013). Seven of the eight studies collected and analyzed psychological metrics, and all
seven of those reported significant changes after the intervention (Andersen et al., 2016;
Bouchard et al., 2012; Dziembowska et al., 2016; Kudo et al., 2014; McCraty et al., 2009;
Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2013). In terms
of performance, all five studies that assessed performance also reported a significant
improvement following the intervention (Andersen et al., 2016; Bouchard et al., 2012; Prinsloo
et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b; Rusciano et al., 2017; Sherlin et al.,
2013).

While biofeedback was also largely effective in the convenience samples, the results are
less comprehensive. Eight of the nine studies collected and analyzed physiological data, and six
of them (75%) reported a significant difference among at least one measure (Astor et al., 2013;
Bradley et al., 2010; Escolano et al., 2014; Kim et al., 2012; Sherlin et al., 2010; Whited et al.,
2014). Six of the nine studies also collected and reported psychological metrics, but only four
total (67%) reported any significant differences (Bradley et al., 2010; Cohen et al., 2016;
Henriques et al., 2011; Sherlin et al., 2010). Although seven of the nine studies collected and
analyzed data representative of performance, only four of those (57%) reported significant
differences (Astor et al., 2013; Cohen et al., 2016; Escolano et al., 2014; Kim et al., 2012).

The difference observed here is stark, and has substantial implications for the application
of biofeedback interventions. Inconsistencies among results in lab-based settings using
convenience samples could imply that this setting is not truly stressful enough, or it could imply
that individuals recruited through convenience sampling do not have high enough levels of stress
to detect a difference following a stress management intervention. If a user’s stress level is
already quite low to begin with, this sort of intervention may not have as much value. But when
biofeedback is administered to individuals who operate under stressful conditions on a regular basis and who require appropriate stress management techniques to perform their job optimally, the intervention is extremely effective. Furthermore, this effectiveness is pervasive and can be observed across measures of physiology, subjective perceptions and cognitive state, as well as performance measures.

Consistency and Standardization in Data Collection and Measurement

Of equal importance when collecting and feeding back physiological data in near-real time is the equipment used to collected biopotential signals, and the software to process and transform that data in a sensitive and accurate fashion (Laborde, Mosley, & Thayer, 2017). There was very little overlap in regards to equipment used. To collect HRV and HR data, the Freeze-Framer/emWave and StressEraser devices were utilized the most, followed by chest-worn HR belts. Despite the frequent use of chest-worn HR belts, the specific models of belts were all unique. EEG data collection was also accomplished using a variety of devices. This has implications for the quality of data due to a broad variety of acquisition settings across different hardware. The rate at which this technology is advancing could make it difficult to standardize the specific equipment used in these studies, but to obtain transparency and reproducibility, the scientific community using these approaches should strive to at least use the same acquisition settings and detection thresholds for universal metrics, such as HRV and EEG.

The physiological measures themselves were largely comparable across studies, with many overlapping HRV components reported. There was, however, no overlap among EEG components reported as significantly different across studies. This may call into question the interpretations of EEG findings within this domain. Although none of the improvements in
specific bandwidths were corroborated by other studies in this review, the absence of overlap observed could be the byproduct of a lack of standardized approaches, settings, experimental designs, and populations recruited. Because these differences existed among every study included, the same results cannot be expected to be replicated. Something more standard and specific, such as time-domain components of HRV (including SDNN, RMSSD, and pNN50), may be better served to make more generalized and universal claims.

Similarly, there was almost no consistency across psychological measures and the instruments administered to approximate those measures. Stress was a prominent and frequently evaluated construct, but the self-report instruments employed to measure stress varied widely. Overall, there were 23 unique scales or inventories used across the 14 studies that evaluated psychological constructs, eliminating the possibility of direct comparison of psychological assessment.

Beyond addressing physiological and psychological components of stress and stress management, a major focus of this review is the practical concern of how stress and stress management can influence task performance. For the purposes of this discussion, I will consider performance metrics and the experimental task insofar as they relate to the population evaluated. Of the seven studies recruiting convenience samples that also evaluated performance, five of those evaluated RT and/or errors made (Astor et al., 2013; Cohen et al., 2016; Escolano et al., 2014; Raaijmakers et al., 2013; Sherlin et al., 2010), which speaks to relative consistency across studies. However, these metrics were assessed using a variety of tasks. Two studies enlisted the a version of the Stroop task to evaluate performance (Escolano et al., 2014; Sherlin et al., 2010), while the remaining studies used either additional general cognitive batteries (Escolano et al., 2014; Raaijmakers et al., 2013) or tasks developed specifically for their research questions.
(Astor et al., 2013; Bradley et al., 2010; Cohen et al., 2016; Kim et al., 2012). Using consistent performance metrics contributes to greater comparability across convenience samples of students and healthy adults, but using the same task would ensure even greater comparability. Although four of these seven studies did report significant improvement according to at least one performance metric (Astor et al., 2013; Cohen et al., 2016; Escolano et al., 2014; Kim et al., 2012), the variety across assessment tools used limits the generalizability of these results and reveals discrepancies in how performance is evaluated in the lab setting.

The studies evaluating performance within deliberately selected populations did use a wide variety of assessment approaches, but the majority of them were domain- and job-specific, and therefore translatable to performance as a soldier (Bouchard et al., 2012), a professional baseball player (Sherlin et al., 2013), a police officer (Andersen & Gustafsberg, 2016), and an elite professional soccer player (Rusciano et al., 2017). Performance among senior managers was evaluated using the Stroop Color-Word Test (Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b). It is worth noting the significant improvement in performance observed in all of these studies, despite the variety in populations recruited and assessment tools used. In fact, this speaks to the value in tailoring the context and the evaluation modality for performance among specialized populations.

Effects of Intervention Frequency and Study Duration on Collected Measures

The frequency of experimental sessions administering the biofeedback intervention was another major component that varied across studies. Most studies reported significant improvements within some measures across all types of data collected and analyzed, regardless of the duration of the study and the frequency of the intervention. For example, Escolano and
colleagues collected and analyzed physiological and performance data, and reported at least one component within each of those types of data to have significantly changed in the experimental group following the intervention (Escolano et al., 2014). Notably, 80% of studies administering the intervention only once (Astor et al., 2013; Cohen et al., 2016; Escolano et al., 2014; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b) saw improvements across all types of data collected and analyzed. Of the studies administering the intervention between 3 and 10 times, 60% likewise saw widespread improvements (Bouchard et al., 2012; Dziembowska et al., 2016; Kim et al., 2012). Finally 67% of those administering the intervention more than 15 times (Kudo et al., 2014; McCraty et al., 2009; Rusciano et al., 2017; Sherlin et al., 2013) saw improvements across all types of data collected and analyzed. This supports the relative effectiveness of biofeedback as an intervention to address stress management, regardless of the frequency of intervention delivery.

When we consider the types of populations with the most widespread improvements in relation to the number of sessions administered, a more nuanced trend emerges. Overall, 25% of those administered in one session that saw widespread improvements were deliberately chosen samples (Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al., 2013b); 67% of those administered between 3 and 10 times were deliberately chosen samples (Bouchard et al., 2012; Dziembowska et al., 2016); and 100% of those administered more than 15 times were deliberately chosen samples (Kudo et al., 2014; McCraty et al., 2009; Rusciano et al., 2017; Sherlin et al., 2013). This observation suggests that in more specialized populations, a more long-term intervention is associated with greater improvements across measures. Also worth noting is that the only studies reporting a lack of significant improvement in one or more types of data collected had recruited convenience samples (Bradley et al., 2010; Henriques et al., 2011;
Raaijmakers et al., 2013; Sherlin et al., 2010; Whited et al., 2014), regardless of the number of sessions in which the biofeedback intervention was delivered.

Practical Implications

Overall, these results seem to support the idea that contextualizing and personalizing a long-term biofeedback intervention for specialized populations would be the most efficacious approach. Despite the improvements observed across populations enrolled, assessments used, and frequency and duration of the intervention, it is clear that deliberately recruited populations experiencing high levels of stress benefit most when the intervention is equally deliberately designed. These claims cannot be made about convenience samples evaluated in the lab setting.

HRV biofeedback training has an established history of effectiveness across many populations and uses (Lehrer, 2013a), as well as a promising future of additional uses including stress management (Lehrer, 2007; Lehrer & Vaschillo, 2008). Notably, throughout the studies reported in this review, HRV biofeedback was used to present the HRV waveform to users the most out of all delivery modalities. In practical terms, it is the most standardized approach with the strongest record of effectiveness, as well as the most profound mechanistic understandings (Eddie et al., 2015; Lehrer, 2013b; Lehrer et al., 2013). Thus, visual biofeedback relying on a HRV waveform display and HRV biofeedback training principles could be a viable path forward to introduce greater standardization in the field, while still affording the flexibility of tailoring variables such as frequency of intervention, duration of sessions, and performance measures in order to obtain results specific to a given population.

Though performance evaluations and measures should vary according to the population enrolled, standardization of measures should extend to physiological and psychological...
assessments as well. The State Trait Anxiety Inventory for Adults was widely used throughout the studies reviewed here, and remains to be widely used for many studies evaluating stress. This instrument is also available as a 6-item, psychometrically-validated short form (Marteau & Bekker, 1992). Certain contexts and research questions would also require additional self-report measures, but maintaining one common measure would allow greater generalizability or comparison across the literature. In terms of physiological variables, EEG and HRV are typically considered to be sensitive, responsive, and robust indicators of stress states (Sharma & Gedeon, 2012). Other work has shown the positive association between taskload and cognitive workload, measured by HR, HRV, and respiratory rate, on a computer-based task, supportive of using advanced wearable sensors to detect stress and workload (Nixon & Charles, 2017). The research question of interest would likely determine specific EEG and HRV components analyzed (Laborde et al., 2017). Using these signals for analysis would thus ensure a level of reproducibility and comparability to past and future work, but would also allow the opportunity to adapt the analytical approach to the context at hand.

Additional Findings and Considerations

My systematic review also revealed that the display of biofeedback was variable across studies. This is an important finding because to optimize the use of biofeedback, it is necessary to have an appropriate display structure and integration with the task. The biofeedback modality used by 100% of researchers in this review was visual, and more than half of the displays (53%) presented physiological information as the HRV waveform (Figure 2.2). Of the studies using HRV waveform as the primary visual component of the biofeedback intervention, 89% reported significant differences in physiology (Bradley et al., 2010; Dziembowska et al., 2016; Kudo et
al., 2014; McCraty et al., 2009; Prinsloo et al., 2011; Prinsloo et al., 2013a; Prinsloo et al.,
2013b; Rusciano et al., 2017; Sherlin et al., 2010; Whited et al., 2014) and 88% reported
significant differences in psychological metrics (Bradley et al., 2010; Dziembowska et al., 2016;
Henriques et al., 2011; Kudo et al., 2014; McCraty et al., 2009; Prinsloo et al., 2011; Prinsloo et
al., 2013a; Prinsloo et al., 2013b; Sherlin et al., 2010) following the intervention. Only 50% of
those using HRV waveform information that evaluated performance detected significant
differences in performance as a result of the intervention (Prinsloo et al., 2011; Prinsloo et al.,
2013a; Prinsloo et al., 2013b; Rusciano et al., 2017). This difference among the percentage of
change observed in performance compared to physiological and psychological measures could
call into question the purported and previously observed curvilinear relationship between stress
and performance (Allen, 1983).

Aside from the HRV waveform display of information, the remaining modalities were
completely unique. This is important to consider because the method and display in which
physiological changes are presented to the user could have a meaningful impact on the
effectiveness of the intervention. Well-designed information displays can facilitate cognition and
performance, but when designed poorly they can have the opposite effect, inducing cognitive
decline and poorer performance (Kontogiannis & Kossiavelou, 1999). Essentially, for the
intervention (a stimulus) to be effective, the human user must access and interpret it (the
perception) readily, and act on it (the response) accordingly. To optimize this stimulus-response
pattern, the stimulus presentation matters. Although visual cognition has a limited capacity,
information presentation can complement and enhance that capacity (Alvarez, 2011;
Summerfield & Egner, 2009). Display components including location (Wickens & Carswell,
1995), format (Atkins, Wood, & Rutgers, 2002; Shah & Carpenter, 1995), icons (Talcott,
Bennett, Martinez, Shattuck, & Stansifer, 2007), and color (Wickens & Andre, 1990) of information objects all have an impact on information integration and decision-making, warranting a closer consideration of and approach to how information is presented. Recent work has confirmed that providing task-relevant information during cognitively demanding tasks can improve performance and problem-solving capacities (Dadashi, Golightly, & Sharples, 2017). If a major component of the task is the ability to self-regulate and manage stress levels, physiological information displayed through a biofeedback modality could be equally valuable as performance feedback.

Conclusion

The findings from this systematic review reveal that biofeedback interventions for stress management have positive effects, and are particularly effective among deliberately selected populations. A biofeedback intervention, which uses physiological output in an effort to mediate real-time internal states, could potentially aid individuals by regaining an allostatic balance conducive to functional cognition and performance. Considering the neurovisceral integration model (Thayer et al., 2012, 2009), results from this review support that restoring one’s physiological state by focusing on HRV, which was demonstrated to be the most common form of self-regulation training, could affect cognitive resources in a way that enhances performance relying on executive functioning and attention processing. This effect is especially pronounced in high-risk professions. If we consider the curvilinear relationship between stress and performance (Figure 2.1) that has been proposed and supported in the literature, this conclusion can be readily understood. Individuals working in high-risk professions are likely operating in the realm of distress, sometimes past the point of optimal stress. Incorporating biofeedback as a stress
management approach is an effective way to reposition oneself in the realm of eustress via behavioral adaptation. Support for this theory was not found within studies recruiting convenience samples, indicative that this population, mostly comprised of college students and healthy adults, may operate at a lower level of stress that experimental stressors cannot as easily elevate. In this case, biofeedback would reasonably be less effective, because there is a lower stress level to manage at the outset.

The pool of literature acquired had some limitations. The relatively limited number of empirical studies that met my inclusion criteria points to a dearth in research specifically on using biofeedback for the purposes of stress management. The potential implications of using biofeedback as a stress management tool are substantial, but the broad variety in approaches and in results are indicative of the limited consistency across studies. Yet based on the results, of studies recruiting specialized populations, stress reduction and performance enhancement may be more likely to be achieved when the task and metrics are specialized according to the population. With the overall intent to use biofeedback as a training intervention in these populations specifically, users would ideally learn to maintain a healthy physiological balance without continued reliance on the biofeedback device in the long term (Schwartz, 2010). Longitudinal follow-up is thus imperative to confirm the long-lasting effects of biofeedback we wish to instill, yet only one of the 17 studies included reported follow-up data of 90 days post-intervention (McCray, Atkinson, Lipsenthal, et al., 2009), a relatively short period to observe a long-term effect. Their own conclusions emphasize the need for a longer follow-up period, citing that chronically stressed individuals, such as those in high-risk professions, require six to nine months for full recovery, especially considering hormonal stress markers such as cortisol and dehydroepiandrosterone, evaluated in their study (McCray, Atkinson, Lipsenthal, et al., 2009).
Only one other study measured cortisol (Bouchard et al., 2012), but authors make no mention of following up beyond the duration of the study.

Despite the limitations, results of this systematic review are promising. While the acute stress response can be adaptive and appropriate in certain contexts, its inappropriate management and continuous exposure has proven to be detrimental to high-risk practitioners (Arora, Aggarwal, et al., 2010; Arora, Sevdalis, Nestel, et al., 2010; Joseph et al., 2016; LeBlanc, 2009; Mazur et al., 2014; Mazur, Mosaly, Hoyle, Jones, & Marks, 2013; Moorthy, Munz, Dosis, Bann, & Darzi, 2003; Pluyter, Buzink, Rutkowski, & Jakimowicz, 2010; Weigl et al., 2016; Wetzel et al., 2006; Yurko, Scerbo, Prabhu, Acker, & Stefanidis, 2010), who are prone to experience frequent exposures to acute and chronic stressors with little recovery time available. The issue of chronic stress and its effects on health, quality of life, and professional performance is also of primary concern (Balch et al., 2009; Rosenstein, 2012). When stress levels are maintained within an optimal range, however, growth and/or positive performance are facilitated (Allen, 1983; Vine et al., 2013). Ensuring optimal perceptions of stress among practitioner professions is integral, and this population has a substantial potential to benefit from biofeedback administration. The findings from this review illustrate that the systematic evaluation, application, and influence of such technology among non-patients, however, is understudied. Collectively, the results from these studies convey beneficial outcomes across physiological, psychological, and performance measures in various high-performing and acutely and chronically stressed populations, but the limited number of studies reviewed point to a need for more widespread implementation and evaluation. The diversity in methods also suggests a potential benefit for standardization among general populations, and specialization among specific populations.
The benefits associated with biofeedback interventions are notable according to this review of the literature, but one question still looms: How is coping accomplished during times of high stress? While individuals learn to cope in personal and professional capacities as a matter of necessity, a real-time intervention such as biofeedback affords the opportunity to suggest coping at specific points in time. The purpose of the next phase of my dissertation work was to evaluate whether using sensors and physiological information streaming in in real time could inform appropriate physiologically-based times to direct adaptive coping behaviors.

Table 2.1, Part 1

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study population</th>
<th>Sample (Expr/Ctrl)</th>
<th>Mean age (stdev)</th>
<th>Modality: Visual</th>
<th>Modality: Auditory</th>
<th>Physio measures</th>
<th>Physio equipment</th>
<th>Physio results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evetovich et al (2007)</td>
<td>College students</td>
<td>12</td>
<td>22 (1.1)</td>
<td>MMG signal</td>
<td>N/A</td>
<td>MMG, EMG</td>
<td>BioPac</td>
<td>Expr vs. Ctrl: ↓ MMG amplitude, EMG amplitude</td>
</tr>
<tr>
<td>McCratty et al (2009)</td>
<td>Corrected officers</td>
<td>75 (43/32)</td>
<td></td>
<td>HRV waveform (coherence)</td>
<td>N/A</td>
<td>Cortisol, DHEA, S-IgA cholesterol, triglycerides, fasting glucose levels, HR, HRV, BP</td>
<td>Freeze-Framer/emWave</td>
<td>Pre-post Expr: ↓ total cholesterol, LDL, total/HDL ratio, glucose levels, DHEA, mean arterial pressure, systolic BP, diastolic BP, mean HR, ↑ LF/HF ratio of HRV; Pre-post Ctrl: ↓ total cholesterol, HDL, glucose levels</td>
</tr>
<tr>
<td>Bradley et al (2010)</td>
<td>High school students</td>
<td>136 (77/59)</td>
<td>15.3 (0.44)</td>
<td>HRV waveform (coherence)</td>
<td>N/A</td>
<td>HRV</td>
<td>Freeze-Framer/emWave</td>
<td>Pre-post Expr vs. Ctrl: ↓ HR, ↑ SDNN, HF, LF, IC</td>
</tr>
<tr>
<td>Sherlin et al (2010)</td>
<td>Healthy adults</td>
<td>43</td>
<td>33.2 (8.77)</td>
<td>HRV waveform (coherence)</td>
<td>N/A</td>
<td>QEEG</td>
<td>StressEraser</td>
<td>Pre-post Expr: ↑ overall alpha; Pre-post Ctrl: ↓ overall alpha</td>
</tr>
<tr>
<td>Henriques et al (2011)</td>
<td>College students</td>
<td>35</td>
<td></td>
<td>HRV waveform (coherence)</td>
<td>N/A</td>
<td>HRV</td>
<td>emWave</td>
<td>Pre-post Expr: no significant differences</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-------------</td>
<td>------------------------</td>
<td>-----</td>
<td>-------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Prinsloo et al (2011, 2013a, 2013b)</td>
<td>Soldier</td>
<td>41 (21/20)</td>
<td>24.9 (5.55)</td>
<td>Red texture partially obscuring user’s field of view. Increased arousal represented by an increasingly obstructed view</td>
<td>Sound of a pumping heart increasing in frequency and loudness with increased arousal</td>
<td>Daily salivary cortisol, HR</td>
<td>Salivette, Polar wireless transmitter belt</td>
<td>Expr vs. Ctrl: ↓ cortisol, HR</td>
</tr>
<tr>
<td>Bouchard et al (2012)</td>
<td>High school students</td>
<td>46 (21/25)</td>
<td>Not reported</td>
<td>ECG tachogram, anxiety profile</td>
<td>N/A</td>
<td>HRV</td>
<td>QECG-3</td>
<td>Pre-post Expr: ↑ R-R intervals</td>
</tr>
<tr>
<td>Kim et al (2012)</td>
<td>College students</td>
<td>Study 1: 36 (19/17); Study 2: 68 (44/24)</td>
<td>Study 1: 23.39 (range: 8); Study 2: 22.06 (range: 9)</td>
<td>Color-coded arousal meter; corresponding game elements changing color to reflect the state of arousal, located directly in attentional field</td>
<td>N/A</td>
<td>HR (arousal level)</td>
<td>Chest-worn HR belt</td>
<td>Study 1 Expr vs. Ctrl: not reported; Study 2 Expr vs. Ctrl: ↓ arousal level</td>
</tr>
<tr>
<td>Astor et al (2013)</td>
<td>Professional male baseball players</td>
<td>5</td>
<td>19.2 (0.45)</td>
<td>Dual Drive Xtreme: car presented in a race, moves at a speed proportional to user success; OR Particle Editor: shapes or lines in movement along with music when parameters are met</td>
<td>Music</td>
<td>QEEG</td>
<td>Electro-Cap International</td>
<td>Pre-post: ↑ central beta (4/5), frontal beta (3/5), parietal beta (3/5)</td>
</tr>
<tr>
<td>Sherlin et al (2013)</td>
<td>College students</td>
<td>28 (16/12)</td>
<td>22 (range: 8)</td>
<td>Across 6 tasks: 3 displayed SCL signal, 2 displayed HRV signal, 1 displayed a combination of HRV and SCL signals</td>
<td>N/A</td>
<td>HRV, SCL, EEG alpha asymmetry</td>
<td>ActiveTwo</td>
<td>Pre-post Expr vs. Ctrl: No significant differences</td>
</tr>
<tr>
<td>Study Authors</td>
<td>Study Sample</td>
<td>Study Group</td>
<td>Expression Group</td>
<td>Control Group</td>
<td>Method</td>
<td>Measurement</td>
<td>Expressions/Results</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>--------</td>
<td>-------------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td>Escolano et al (2014)</td>
<td>College students</td>
<td>19 (10/9)</td>
<td><strong>Expr</strong>: 25.8 (4.1)</td>
<td><strong>Ctrl</strong>: 24.3 (3.7)</td>
<td>N/A</td>
<td>EEG</td>
<td>Guger Technologies</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A square with changing saturation colors; physiological values reflecting baseline appear gray, values above baseline appear red and increase in saturation as values increase, values below baseline appear blue and increase in saturation as values decrease</td>
<td></td>
<td></td>
<td></td>
<td>Pre-post Expr: ↑ UA power, LA2</td>
<td></td>
</tr>
<tr>
<td>Kudo et al (2014)</td>
<td>Post-partum mothers</td>
<td>55 (25/30)</td>
<td><strong>Expr</strong>: 30.5 (5.7)</td>
<td><strong>Ctrl</strong>: 33.4 (6.6)</td>
<td>N/A</td>
<td>HRV</td>
<td>StressEraser</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HRV waveform (coherence)</td>
<td></td>
<td></td>
<td></td>
<td>Pre-post Expr vs. Ctrl: ↓ HR, ↑ SDNN</td>
<td></td>
</tr>
<tr>
<td>Whiled et al (2014)</td>
<td>Undergraduate and graduate students</td>
<td>28 (15/13)</td>
<td><strong>Expr</strong>: 22.29 (3.6)</td>
<td><strong>Ctrl</strong>: 23.15 (4.06)</td>
<td>N/A</td>
<td>HR, HRV</td>
<td>Polar RS800CS, emWave</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HR, HRV power spectrum, HRV waveform (coherence)</td>
<td></td>
<td></td>
<td></td>
<td>Pre-post Expr vs. Ctrl: ↑ pNN50</td>
<td></td>
</tr>
<tr>
<td>van Dijk et al (2015)</td>
<td>Healthy adults</td>
<td>74</td>
<td>27 (range: 49)</td>
<td></td>
<td>N/A</td>
<td>HR</td>
<td>Kendall H124SG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trending HR over last 3 minutes, current HR in BPM</td>
<td></td>
<td></td>
<td></td>
<td>Expr HR vs. Ctrl: ↑ correlation between HR and momentary/retrospective stress</td>
<td></td>
</tr>
<tr>
<td>Zauszniewski et al (2015)</td>
<td>Grandmothers raising grandchildren</td>
<td>20</td>
<td>58 (range: 26)</td>
<td></td>
<td>N/A</td>
<td>HRV</td>
<td>StressEraser</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HRV waveform (coherence)</td>
<td></td>
<td></td>
<td></td>
<td>(-) correlation between perceived stress and coherence, negative emotions and coherence, and depressive cognitions and coherence</td>
<td></td>
</tr>
<tr>
<td>Andersen &amp; Gustafsson (2016)</td>
<td>Police officers</td>
<td>12 (6/6)</td>
<td>31.5 (range: 7)</td>
<td></td>
<td>N/A</td>
<td>HR, BP</td>
<td>Zephyr BioHarness, Omnisense ambulatory blood pressure wrist cuffs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No details</td>
<td></td>
<td></td>
<td></td>
<td>Pre-post Expr vs. Ctrl: ↓ HR max between baseline and Scenario 1, ↓ change in HR from baseline to Scenario 1</td>
<td></td>
</tr>
<tr>
<td>Cohen et al (2016)</td>
<td>Healthy adults</td>
<td>29</td>
<td>25.5 (4.67)</td>
<td></td>
<td>N/A</td>
<td>HR, HRV</td>
<td>Zephyr HxM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Performance prediction bar graph, HR bar graph, error chance prediction bar graphs</td>
<td></td>
<td></td>
<td></td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Dziembowska et al (2016)</td>
<td>Young male athletes</td>
<td>41 (20/21)</td>
<td>18.34 (1.36)</td>
<td></td>
<td>N/A</td>
<td>EEG alpha asymmetry, HRV</td>
<td>emWave, Electro-Cap International</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HRV waveform (coherence)</td>
<td></td>
<td></td>
<td></td>
<td>Pre-post Expr: ↑ HRV spectral power, LF HRV, IC, theta and alpha spectral power, alpha asymmetry, ↓ HF HRV</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Psych measures</td>
<td>Psych instruments</td>
<td>Psych results</td>
<td>Performance measures</td>
<td>Performance instruments</td>
<td>Performance results</td>
<td>Session</td>
<td>Frequency</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>-------------------------</td>
<td>----------------------</td>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>Rusciano et al (2017)</td>
<td>Profession: male soccer players</td>
<td>Expr: Sessions 1-3: HRV waveform (coherence); Sessions 4-9: HRV waveform (coherence); Sessions 10-15: HRV waveform (coherence)</td>
<td>Ctrl: Sessions 1-3: SCL, RR, BVP, HR, HRV, surface EMG, skin temperature</td>
<td>Expr vs. Ctrl: ↓ SCL, HR, RR; Pre-post Expr: ↑ LF</td>
<td>Nexus 10 Mark II</td>
<td>15-min Heart Lock-ins, 3 FreezeFrames per week; 30-sec FreezeFrames/hr</td>
<td>Daily</td>
<td>3 months</td>
</tr>
<tr>
<td>Evetovich et al (2007)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Number of repetitions</td>
<td>Seated preacher curl exercise</td>
<td>Expr vs. Ctrl: No significant differences</td>
<td>Until fatigued</td>
<td>1 session every 2 days (5 total per participant)</td>
</tr>
<tr>
<td>McCraty et al (2009)</td>
<td>Health risk, Type A and coronary-prone behavior, psychological symptoms, psychological and workplace elements</td>
<td>Personal Wellness Profile, Jenkins Activity Survey, Brief Symptom Inventory, Personal and Organizational Quality Assessment</td>
<td>Pre-post Expr: ↓ anger, fatigue, hostility, interpersonal sensitivity, paranoid ideation, psychological distress, Type A behavior, speed and impatience, ↑ productivity, motivation, goal clarity, perceived manager support, gratitude, positive outlook; Pre-post Ctrl: ↓ work attitude, confidence in organization, ↑ depression</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>5 months</td>
<td></td>
</tr>
<tr>
<td>Bradley et al (2010)</td>
<td>Test anxiety (global, worry, emotionality)</td>
<td>Test Anxiety Inventory (8 of 16 items)</td>
<td>Pre-post Expr vs. Ctrl: ↓ overall test anxiety (global, worry, emotionality)</td>
<td>Standardized testing scores</td>
<td>Californi a High School Exit Exam, California Standard s Test</td>
<td>Pre-post Expr vs. Ctrl: No significant differences</td>
<td>TestEdge</td>
<td>2 times per week</td>
</tr>
<tr>
<td>Study</td>
<td>State and Trait Anxiety</td>
<td>Interview-State Form</td>
<td>Pre-post Expr vs. Ctrl: ↓ state anxiety</td>
<td>Incorrect versus correct response, reaction time, missed responses</td>
<td>Modified Stroop task</td>
<td>Pre-post Expr vs. Ctrl: No significant differences</td>
<td>15 minutes</td>
<td>Once</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------</td>
<td>-----------------------</td>
<td>----------------------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------</td>
<td>------------------------------------------------</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td>Henriquez et al (2011)</td>
<td>Mood/a anxiety symptoms, state and trait anxiety, subjective well-being</td>
<td>Mood and Anxiety Symptom Questionnaire, State Trait Anxiety Inventory, Scales of Psychological Well-Being</td>
<td>Pre-post Expr: ↓ general distress: mixed, general distress; anxious, anxious arousal</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prinsloo et al (2011, 2013a, 2013b)</td>
<td>State and trait anxiety, relaxation, sleepiness</td>
<td>State-Trait Anxiety Inventory Forms Y-1 and Y-2, Smith Relaxation States Inventory 3, Visual Analog Scale</td>
<td>Pre-post Expr vs Ctrl: ↑ relaxation, ↓ sleepiness</td>
<td>Reaction time, mistakes made</td>
<td>Stroop task</td>
<td>Pre-post Expr vs. Ctrl: ↑ reaction time; consistency in responses, ↓ mistakes made</td>
<td>10 minutes</td>
<td>Once</td>
</tr>
<tr>
<td>Kim et al (2012)</td>
<td>N/A</td>
<td>N/A</td>
<td>Speech performance</td>
<td>Korean Broadcasting System speech performance test</td>
<td>Pre-post Expr: ↑ speech performance</td>
<td></td>
<td></td>
<td>3 total sessions</td>
</tr>
<tr>
<td>Astor et al (2013)</td>
<td>Cognitve reappraisal and suppression</td>
<td>Emotion Regulation Questionnaire</td>
<td>Not reported</td>
<td>Percentage of correct decisions</td>
<td>Auction Game decision performance</td>
<td>Study 1 Expr vs. Ctrl: ↑ performance; Study 2 Expr vs. Ctrl: ↑ performance</td>
<td>25 minutes</td>
<td>Once</td>
</tr>
<tr>
<td>Sherlin et al (2013)</td>
<td>Focus index, stress index, speed index</td>
<td>NeuroPerformance Profile</td>
<td>Pre-post: ↑ sleep, focus, feeling of being relaxed, all indices of NeuroPerformance Profile</td>
<td>Response time, response time variability, errors of omission, errors of commission</td>
<td>QIKtest continuous performance test</td>
<td>Pre-post: ↓ errors of commission</td>
<td>20-35 minutes</td>
<td>Averages of 3 session per week (15 total per participant)</td>
</tr>
<tr>
<td>Study</td>
<td>Affect Measures</td>
<td>Pre-post Expr vs. Ctrl:</td>
<td>Reaction Time, Error Percentages</td>
<td>N-back Task, Mental Rotation Task</td>
<td>Pre-post Expr vs. Ctrl:</td>
<td>Reaction Time, Error Percentages</td>
<td>N-back Task, Mental Rotation Task</td>
<td>Duration</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------</td>
<td>-------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
<td>-------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Raaijmakers et al (2013)</td>
<td>Positive and Negative Affect Scale</td>
<td>No significant differences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30 minutes</td>
</tr>
<tr>
<td>Escolano et al (2014)</td>
<td>N/A</td>
<td>N/A</td>
<td>Errors, elapsed time, recognized words, interference, correct responses, reaction time</td>
<td>Paced Auditory Serial Addition Task, Rey Auditory Verbal Learning Test, Trail Making Test, Stroop Color-Word Test, Mental Rotation Task</td>
<td>Pre-post Expr vs. Ctrl:</td>
<td>↑ part B of TMT; Pre-post Expr: ↓ errors and time elapsed in PASAT, reaction time in MRT, ↑ recognized words in RAVLT, correct responses in MRT, part A of TMT</td>
<td>Pre-post Expr vs. Ctrl:</td>
<td>25 minutes</td>
</tr>
<tr>
<td>Kudo et al (2014)</td>
<td>Postnatal depression</td>
<td>↓ EPDS score, anxiety, difficulty sleeping, sad and miserable thoughts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>To a certain score (untimed)</td>
</tr>
<tr>
<td>Whitfield et al (2014)</td>
<td>Distress, stress, state affect</td>
<td>No significant differences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32 minutes</td>
</tr>
<tr>
<td>van Dijk et al (2015)</td>
<td>Momentary stress, retrospective stress, neuroticism, anxiety sensitivity</td>
<td>Ctrl: (+) correlation between neuroticism and momentary/retrospective stress; Expr_Stress: (+) correlation between anxiety sensitivity and momentary/retrospective stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20 minutes</td>
</tr>
<tr>
<td>Study</td>
<td>Perceived Stress Scale, Emotional Symptom Checklist, Depressive Cognition Scale</td>
<td>(-) correlation between perceived stress and coherence, negative emotions and coherence, and depressive cognitions and coherence</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>10-15 minutes</td>
<td>At least 28 sessions</td>
<td>3 months</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----------------</td>
<td>----------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Andersen &amp; Gustafsberg (2016)</td>
<td>Stress, Skills, Ability to manage scenarios, Situational awareness</td>
<td>Pre-post Expr vs. Ctrl: ↓ stress, ↑ confidence in skills/ability to manage critical incidents</td>
<td>Critical</td>
<td>Senior</td>
<td>Pre-post Expr vs. Ctrl: ↑ critical task completion, situational awareness, decision making, overall performance</td>
<td>International performance resilience and efficiency program (iPREP)</td>
<td>N/A</td>
<td>6 days</td>
</tr>
<tr>
<td>Cohen et al (2016)</td>
<td>Likert-type scale, Level-of-Information Processing Scale, System Usability Scale</td>
<td>Expr_HR vs. Expr_noHR: ↓ perceived usability of the feedback</td>
<td>Total score for task, number of errors made</td>
<td>Simulate d fire management task</td>
<td>Expr vs. Ctrl: ↑ relative performance score</td>
<td></td>
<td>1 session 1 day</td>
<td></td>
</tr>
<tr>
<td>Dziembowska et al (2016)</td>
<td>State-Trait Anxiety Inventory A-State Scale, Rosenberg Self Esteem Scale</td>
<td>Pre-post Expr: ↓ state anxiety</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>10 sessions 3 weeks</td>
<td></td>
</tr>
<tr>
<td>Rusciano et al (2017)</td>
<td></td>
<td>Accuracy, reaction time, injury prevention</td>
<td>Stroop, visual search task</td>
<td>Pre-post Expr vs. Ctrl: ↑ accuracy in Stroop task, presence at training days, RT in visual search task, absences at training days</td>
<td>15-30 minutes</td>
<td>15 sessions 4 months</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3
Timing of coping instruction presentation for real-time acute stress management: Potential implications for improved surgical performance

Abstract

Individual performance on complex healthcare tasks can be influenced by acutely stressful situations. Real-time biofeedback using passive physiological monitoring may help to better understand an individual’s progression toward acute stress-induced performance decrement. Providing biofeedback at an appropriate time may provide learners within an indicator that their current performance is susceptible to a decrement, and offer the opportunity to intervene. I explored the presentation timing of coping instructions during an acutely stressful task. In this pilot study, I recorded and analyzed electrocardiography data surrounding coping instruction presentation on various time schedules while participants played a first-person shooter computer game. Around times of significantly elevated heart rate, an indicator of acute stress, presenting a coping instruction tended to result in an increase in heart rate variability (HRV) following its presentation, with a more marked effect in high stress conditions; not presenting a coping instruction at this time tended to result in a decrease in HRV in high stress conditions, and no change in low stress conditions. HRV following instruction presentation tended to increase in both high and low stress conditions when the instruction was presented at times of elevated heart rate; there was very little change in HRV when instruction presentation was not bound to physiology. Performance data showed that better performance was associated with greater adherence to coping instructions, compared to when zero instructions were
followed. Implications for healthcare are significant, as acute stress is constant and it is necessary for providers to maintain a high level of performance.

Introduction

The experience of acute stress is a natural and evolutionarily adaptive physiological response (Allen, 1983), and is modulated by the cognitive perception of the resources and demands of the situation. The theory of cognitive reappraisal suggests that when we perceive our available resources to be sufficient to meet the demands posed by the stressor, a state of challenge ensues (Lazarus & Folkman, 1984). This state is defined by the arousal of the sympathetic nervous system (SNS), which is characterized by an increase in heart rate, pupil diameter, respiratory rate, and a redirection of blood flow to large peripheral muscle groups (Venables & Fairclough, 2004; Zhai & Barreto, 2006), among other changes. These changes enable us to either fight or flee from the stressor.

However, when the available resources are perceived as insufficient to meet the demands of the stressor, a state of threat ensues (Lazarus & Folkman, 1984). Evidence shows that the distinction between the perception of challenge and the perception of threat has physiological underpinnings which further extend to affect one’s cognitive state (LeBlanc, 2009). While the same SNS response unfolds following the perceptions of challenge and threat, the perception of threat can be distinguished by downstream endocrine changes. The systemic release of cortisol associated with the threat response has a delayed effect, and once it reaches and floods key neural areas, cognitive functioning characteristic of those areas begins to decline. Specifically, the amygdala, hippocampus, and prefrontal cortex are affected, resulting in a decline in processes such as attention, memory recall, and decision-making (LeBlanc, 2009).
Due to the signature endocrine and physiological changes that result from the presence of an acute stressor, measuring these changes to detect the onset of acute stress is highly accurate and reliable. Heart rate variability (HRV) is one specific measurement that is extremely sensitive to acute stress detection (Tanev et al., 2014; Zhai & Barreto, 2006), derived from heart rate (HR) data from electrocardiographic (ECG) acquisition methods. Analysis of HRV can focus on frequency-domain or time-domain calculations, each of which is comprised of distinct features and is determined by distinct mathematical manipulations.

Frequency-domain analysis applies spectral methods to determine the pattern of variance distributed across the data segment as a function of its frequency (Electrophysiology, 1996). Features reported within the frequency-domain most typically include very low frequency, low frequency, high frequency, and the low frequency: high frequency ratio. Time-domain analysis involves detecting R peaks from a segment of raw ECG and performing calculations based on the duration between consecutive R peaks (Electrophysiology, 1996). Time-domain features most notably include the root mean square of the successive differences (RMSSD), standard deviation from normal-to-normal peak (SDNN) and the percentage of consecutive peak durations differing by more than 50 milliseconds (pNN50).

Of the HRV components typically evaluated, analysis using RMSSD is demonstrably highly accurate at classifying the level of emotional arousal over 15-second time intervals (Schaaff & Adam, 2013). This knowledge is valuable in analyzing data for short time windows post-hoc, but is also relevant to understanding the nature of biofeedback and the potential use of biofeedback in mobile environments.

Biofeedback training, often utilizing and displaying HR or HRV information from ECG data, enables users to visualize their own physiological responses and ultimately to manage those
physiological responses. The purpose is to understand one’s own physiological response and learn to control it, with the goal of improved health and performance (Ortiz-Vigon Uriarte, Garcia-Zapirain, & Garcia-Chimeno, 2015). As information is fed back to the user, alterations in cognition, emotion, and behavior affect physiology iteratively (Schwartz, 2010). By monitoring signs of stress in real time, users learn to preemptively acknowledge and manage elevating stress levels (Eddie et al., 2015; Lehrer, 2007; Lehrer & Gevirtz, 2014). Biofeedback is gaining popularity as a stress management tool by providing individualized physiological information in real time to the user, and is by definition immediate and specific, landmark features of influential and lasting performance feedback (Scheeler, Ruhl, & McAfee, 2004).

Inappropriate design or implementation of a biofeedback intervention could lead to interruptions, distracting information, and ambiguous coping instructions (Kontogiannis & Kossiavelou, 1999). To optimize human performance using biofeedback, it is critical that the display of information is clear, concise, and usable. On its own and without adequate training, biofeedback is not inherently usable or assistive in nature, necessitating the investigation of a supplemental coping instruction to present alongside biofeedback to aid in interpretation.

Psychophysiological-based, real-time coping interventions have rarely been investigated, but have proven successful (G. F. Wilson & Russell, 2007). The goal of this pilot study was to evaluate the most effective time to present coping instructions as an intervention on its own, before combining it with biofeedback, during an acutely stressful task (a first-person shooter video game, Counter-Strike: Global Offensive [CSGO]) to build on existing biofeedback literature, and to test the potential for this type of intervention into fast paced work settings. Improvement was defined as a decrease in physiological and cognitive measures reflecting higher stress states, and an increase in task performance.
I hypothesized that receiving a brief coping instruction versus not receiving this instruction at physiologically-relevant times would result in a greater reduction in physiological stress indicators. I also expected that receiving this coping instruction in response to elevations in physiology would result in a greater reduction in stress indicators as compared to receiving the coping instruction on prescribed time schedules, which are inherently agnostic to physiological changes. Furthermore, I hypothesized that performance on the task would be better when more coping instructions were followed, representing greater adherence to the proposed intervention.

Related Literature

The relationship between elevated physiological indicators of acute stress and the associated decline in critical domains of cognitive processing is clear and well-documented (Bohnen, Houx, Nicolson, & Jolles, 1990; Buchanan & Lovallo, 2001; Buchanan, Tranel, & Adolphs, 2006; de Quervain, Roozendaal, Nitsch, McGaugh, & Hock, 2000; LeBlanc, 2009; Lupien, Gillin, & Hauger, 1999). Along with a decline in key cognitive functions, such as attention, memory, and decision-making, excessive or prolonged episodes of acute stress can subsequently lead to impaired task performance (Bohnen et al., 1990). Notably, the relationship between physiology, cognition, and performance has been described in general populations, and more specifically among healthcare practitioners (Arora, Sevdalis, Nestel, et al., 2010; Mazur et al., 2014, 2013; Moorthy et al., 2003; Pluyter et al., 2010; Weigl et al., 2016; Wetzel et al., 2006; Yurko et al., 2010). Thus, these effects associated with the acute stress response are universal phenomena, transferrable across populations. Additionally, such results have been demonstrated during clinical practice (Joseph et al., 2016), in higher-fidelity simulated settings or tasks (Arora, Sevdalis, Aggarwal, et al., 2010; Moorthy et al., 2003; Weigl et al., 2016; Yurko et al., 2010),
and in lab-based settings using medically-oriented computer tasks (Kennedy & Parker, 2017; Mazur et al., 2014, 2013) alike.

One example of a cognitive architecture generalizable across levels of fidelity and experimental settings is the Cognitive Load Theory (CLT) (Sweller, 1988). CLT emphasizes mental workload, the limited capacities of different categories of mental loads, the interactions of these loads during a task, and the reduction in cognitive processing and task performance associated with mental overload (Galy, Cariou, & Mélan, 2012). This framework has been previously applied to medical education, e-learning, clinical diagnosis, and ergonomics settings (Galy & Melan, 2015; Van Merriënboer & Sweller, 2010). The primary assumption of CLT is a limited mental capacity, which can be manipulated by addressing the characteristics of and interactions between the three mental load categories (Van Merriënboer & Sweller, 2010). The three categories of mental loads described in CLT include intrinsic, extraneous, and germane cognitive load.

Intrinsic load is characterized as the load induced by the material or task components themselves, including task difficulty (Sweller, 1988). In the scheme of the present study, we would thus expect intrinsic load to vary with the difficulty level of the condition. Since we expect the higher difficulty version of the task to induce a higher level of perceived stress and cognitive workload, I hypothesize that self-report responses will indicate higher stress and workload in the conditions with highest intrinsic load.

Extraneous load is modulated by external factors, such as the overall situation, time pressure, and environmental noise (Sweller, 1988). Extraneous load is not expected to vary across participants or conditions in the current study (see Equipment section below).
Finally, germane load was originally defined as the load induced by working memory during automation (Sweller, 1988), but has more recently been revisited and more specifically defined as the load induced by the conscious effort of applying strategies to increase task efficiency (Schnotz & Kürschner, 2007). It is this category that is of most interest in the current study reported. My goal is to externally supply a strategy to enable participants to readily cope with the task-associated stressors, and thereby reduce the demands associated with germane load. I propose to do this by relieving participants of the requirement to implement coping strategies on their own. I expect the instructions to be most effective in reducing germane load, and by extension overall load, when they are most salient to the overarching goals—that is, when they are introduced at times of high physiological arousal. Thus I hypothesize that physiological indicators of stress will be lower following the administration of coping instructions at physiologically-relevant times compared to their a) lack of administration as physiologically-relevant times and b) administration at physiologically-arbitrary times. I would interpret that the observation of this trend would be reflective of a reduced germane load.

Psychophysiological analysis will focus on HRV measures, which represent a greater sensitivity to detecting changes in mental state compared to HR (Böhm, Rötting, Schwenk, Grebe, & Mansmann, 2001). HRV is sensitive not only to changes in physical state as has been the traditional understanding, but also to changes in mental state (Taelman & Vandeput, 2011). This finding was demonstrated using 6-minute, lab-based computer tasks with cognitive demands (Taelman & Vandeput, 2011), paralleling the experimental design of the current study in some ways. Furthermore, tasks involving minimal physical effort have demonstrated an ability to detect changes in mental state with HRV analysis alone (Brunken, Plass, & Leutner, 2003).
According to the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology report published in 1996, the minimum time window required to generate reproducible, valid, and accurate HRV data is five minutes for time-domain and frequency-domain features alike (Electrophysiology, 1996). More recent work, however, has shown the reproducibility, validity, and accuracy in analysis of time-domain features for time windows as limited as 10-30 seconds (D. Kim, Seo, Kim, & Jung, 2008; Munoz et al., 2015; Nussinovitch et al., 2011; Salahuddin, Cho, Jeong, & Kim, 2007; Schaaff & Adam, 2013; Schroeder et al., 2004; Thong, Li, McNames, Aboy, & Goldstein, 2003). Of the time-domain measures, RMSSD is most appropriate and consistent in detecting differences in mental stress and arousal (Munoz et al., 2015; Salahuddin et al., 2007; Schaaff & Adam, 2013). HRV analysis will thus specifically focus on RMSSD as an indicator of acute stress.

Performance results from the current study using CSGO may also contribute to the overall framework of CLT, and specifically to our understanding of germane load. If the intervention is adhered to more consistently over the course of the task, we would expect overall performance to improve due to a reduction in germane load required in enlisting coping strategies to apply to the task.

Application Environment

Acute stress is a well-documented challenge for surgeons and may influence surgical performance (Arora, Sevdalis, Nestel, et al., 2010; Healy & Tyrrell, 2011), and has long-established physiological underpinnings (Allen, 1983). Physiological and cognitive effects of SNS activation in the face of acute stress detrimentally impact cognitive processes integral in high performance (Allen, 1983), which could result in preventable errors. Medical error in the
United States has been estimated as the third leading cause of death (Makary & Daniel, 2016) and surgery represents an acute care setting particularly susceptible to errors (Gruen, Jurkovich, McIntyre, Foy, & Maier, 2006). Managing stress is imperative to optimize surgical performance and limit medical errors.

A systematic review of the impact of distractions on surgical performance has shown that equipment and procedural distractions are the most severe according to observational studies, and auditory and mental distractions negatively impact surgical performance according to laboratory studies (Mentis, Chellali, Manser, Cao, & Schwitzberg, 2016). Considering the novelty of the proposed intervention (biofeedback) and its potential to introduce mental distractions if poorly designed, the components and form of this intervention must be systematically evaluated to ensure optimal usability.

Within populations of surgeons, prior general video game experience has been shown to correlate to higher baseline laparoscopic skills (Jalink, Goris, Heineman, Pierie, & Ten Cate Hoedemaker, 2014) and robotic skills simulator performance (Harbin et al., 2016) and decreased time needed to acquire surgical skills among novices (Shane, Pettitt, Morgenthal, & Smith, 2008). Video game “warm-up” has also been demonstrated to improve surgical skills (Jalink, Heineman, Pierie, & ten Cate Hoedemaker, 2015; Rosser, Jr., Gentile, Hanigan, & Danner, 2012). Finally, systematic video game training, using a first-person shooter task similar to the one used in this pilot study, improved surgical performance among a group of surgical novices (Schlickum, Hedman, Enochsson, Kjellin, & Fellander-Tsai, 2009).

Using a first-person shooter game similar to others (e.g. Schlickum et al., 2009), I recruited healthy undergraduates as participants in this study. The current study sought to evaluate the relatively new concept of supplemental coping instructions introduced for acute
stress management. Therefore the current study recruited individuals from the general population and used a generally stressful task.

Current Study

The existing literature surrounding the utility of serious games for training and improving skills among healthcare practitioners is growing (Graafland, 2014; Graafland et al., 2014; Wattanasoontorn, 2013). These serious games are designed for individuals with medical training. Further, many effective serious games in healthcare typically include specific skill sets within medical specialties (Graafland, Schraagen, & Schijven, 2012). The current study sought to evaluate the relatively new concept of supplemental coping instructions introduced for acute stress management. Therefore the current study recruited individuals from the general population and used a generally stressful task.

Researchers using the same first-person shooter game to evaluate the effect of stress on cognitive processing among undergraduate students showed that CSGO induces a threat response ultimately leading to impaired learning and memory (Kindermann, Javor, & Reuter, 2016). Additional work has shown that a highly arousing first-person shooter video game similar to CSGO induced poorer information processing and impaired cognitive performance overall, also among undergraduate students (Maass, Klöpper, Michel, & Lohaus, 2011).

In summary, while there is some support for incorporating psychophysiological-derived coping interventions to improve task performance (G. F. Wilson & Russell, 2007), there is an overall dearth in evidence. The current study aims to generate further evidence to support the benefit of including explicit coping strategies during an acutely stressful task to improve: the
subjective perception of stress and cognitive workload; the physiological indicators of acute stress; and overall task performance.

My specific research questions and corresponding hypotheses are as follows:

- Do the perceived levels of stress and cognitive workload change when the task difficulty increases?
  
  *Hypothesis 1*: Responses to the self-report questionnaires will indicate a higher level of stress and a higher perceived cognitive workload in the more difficult conditions, indicative of a higher level of intrinsic load, with increased task difficulty.

- When HR is substantially elevated, is there a difference between receiving and following a coping instruction at this time, versus not receiving a coping instruction at this time?
  
  *Hypothesis 2*: Physiological measures downstream of times of substantially elevated HR will be representative of a lower level of stress when a coping instruction is presented and followed, compared to when a coping instruction is not presented or not followed. Removing the requirement to devote cognitive resources to coping due to this intervention will decrease the level of germane load.

- Does receiving and following a coping instruction have a different effect when the instruction is physiologically-relevant, versus when it is physiologically-arbitrary?
  
  *Hypothesis 3*: Physiological measures downstream of times when coping instructions are presented will reveal that stress is more effectively reduced when the coping instructions presented are based on changes in physiology, compared
to when coping instructions presented are not based on changes in physiology. Providing an explicit coping strategy when it is needed will be less distracting than providing one when it is not needed, and will decrease the level of germane load.

- Does performance differ as a result of increased intervention adherence?
  - **Hypothesis 4**: Performance will improve when the intervention is adhered to more consistently over the course of the condition and task. Following the instructions of an explicit coping strategy throughout the task will reduce the amount of germane load required.

An earlier version of this research, published in IEEE conference proceedings (Kennedy & Parker, 2016), includes preliminary analysis of physiological data related to Hypothesis 2. The current study greatly expands upon this analysis by introducing novel research questions, novel analytical approaches, and an overall broader context, framework, and interpretations in which to consider the corresponding findings.

**Materials and Methods**

**Participants**

A total of 21 students (11 females) were recruited at a liberal arts college in the southeast United States. For 19 of these participants, CSGO was a novel task; 2 participants (both males) had experience playing this game in the past. A total of 15 participants had experience playing at least one of many similar first-person shooter video games in the past. All 15 participants with prior similar experience had played Call of Duty, and many had also played games such as
Destiny, Battlefield, Halo, Fallout, Borderlands, BioShock, and Battlefront. Participants’ ages ranged from 18 to 23 years. Normal or corrected-to-normal vision was the primary inclusion criteria. The experiments were approved by and conducted in accordance with the guidelines of the Roanoke College Institutional Review Board, including written informed consent.

Equipment

Physiological signals (electroencephalography [EEG], electrocardiography [ECG], and skin temperature) were recorded with AD Instruments, Inc. equipment (Colorado Springs, CO), namely the PowerLab 26T device. Electro-Cap International, Inc. (Eaton, OH) EEG caps transmitted voltage signals from the scalp of participants at sites Fz, Pz, O2, and two mastoid grounds. Three lead-shielded electrode leads attached to disposable ECG electrodes transmitted voltage signals from the chest and wrist of participants. Skin temperature was recorded on participants’ non-dominant hand using a skin temperature sensor. Stimulus presentation times were indicated by a signal sent from an external Cedrus StimTracker (San Pedro, CA) device to a separate monitor, and were also recorded by LabChart 7 software. Signals were presented on an external 17” Dell monitor viewed by the researcher only. Stimuli were presented in line of sight on the internal 15” monitor of a Dell laptop using SuperLab 4.5 (San Pedro, CA). Participants used a Lenovo Yoga 12 laptop computer consisting of an internal 12” monitor to complete the task.

In the current study, experimental protocol and physical organization of the experimental setting were heavily standardized to minimize the influence of extraneous load on cognition and performance across participants. Each participant completed the study individually, in a behavioral booth designed for human subjects’ research, with consideration given to reduce
external noise and maintain a stable room temperature. Physiological sensors used in this study were minimally invasive and had little effect on participants’ overall perception and ability to complete the task.

Procedure

Participants first reviewed and signed the informed consent, and then were equipped with EEG, ECG, and skin temperature sensors. Participants then played the two experimental game modes of CSGO (Hidden Path Entertainment and Valve Corporation, Bellevue, WA) for three minutes each to familiarize themselves with the game and the controls. Participants then reported demographic information and responded to form Y-2 of the State Trait Anxiety Inventory for Adults (STAI) to report the general trait of stress in daily life. Five minutes of baseline data were acquired while the participant relaxed in silence. The researcher calculated average HR over the five minute baseline period, and two standard deviations above the average resting HR.

Participants then played ten minutes of each condition, in randomized order. The message reading “Take a deep breath.” visually appeared according to various time schedules: fixed intervals (Fixed Time; FT), randomly (Variable Time; VT), elicited by an elevation of two standard deviations above resting HR (Physiologically-Triggered; PT), or not at all (Control; CON). In the FT condition, a message was displayed every two minutes, a total of five times. In the VT condition, a message was displayed five times randomly. In the PT condition, the frequency of message presentation depended on how often the individual’s HR exceeded their personalized threshold. Stress level varied according to the game mode played. Weapons Course was considered the low stress mode, in which participants completed the game tutorial. Arms Race was the high stress mode, which consisted of a team-based, first-person shooter game with
high visual-spatial demands and only one winner. The experimental design was thus a 4 (presentation timing) x 2 (stress level) within-subjects design.

Between each condition, participants responded to form Y-1 of the STAI to report their current state of stress as a result of participating in the corresponding condition and the NASA Task Load Index (NASA-TLX) to report perceived cognitive workload associated with each specific condition.

Analysis

Analysis was completed using LabChart Reader by AD Instruments (Colorado Springs, CO), Microsoft Excel, and JMP Pro 12 by SAS Institute, Inc. (Cary, NC). Repeated measures ANOVAs and paired Student’s t-tests were used to compare the means of all data acquired.

HRV was derived from the raw ECG trace by calculating durations between consecutive R-R peaks for the 30 seconds leading up to and the 30 seconds following events of interest. Events of interest were substantial elevations in heart rate, as well as stimulus presentation. HRV measures were calculated on an individual basis surrounding events of elevated HR for times in which a message was displayed (PT conditions) and compared to analogous moments of elevated HR when a message was not displayed (all other conditions). The same method of analysis was used to calculate HRV preceding and following individual stimuli. In this case, stimuli elicited in response to physiology, which are considered to be physiologically-relevant (PT conditions), were compared to stimuli elicited on a prescribed time schedule, which are considered to be physiologically-arbitrary (FT and VT conditions).
RMSSD was the specific time-domain measure selected to calculate HRV for this subset of events. RMSSD is calculated using the following formula, where N is equal to the number of R-R intervals:

\[
RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} ((R - R)_{i+1} - (R - R)_i)^2}
\]

A filtering system was applied to the entire dataset, to calculate RMSSD values corresponding to the 30-second time windows preceding and following peaks of elevated stress, and preceding and following stimulus presentation. This filtering system was applied to arrive at a subset of data points that were most isolated from surrounding physiological or stimulus-related events. The data points remaining therefore represented those with minimal influences that could potentially inappropriately influence the interpretation of results.

I approached analysis of physiological data to answer two separate questions, each with a corresponding hypothesis. The first question is concerned with the changes in RMSSD before and after substantial peaks in HR, and the comparison is between times when a coping instruction was presented (PT conditions) versus times when a coping instruction was not presented (VT, FT, CON; Other conditions). Predicted outcomes are detailed in Hypothesis 2. The question can be summarized as follows: when HR is substantially elevated, is there a difference between receiving and following a coping instruction at this time, versus not receiving a coping instruction at this time?

The second question addressed through this analysis is concerned with changes in RMSSD before and after stimulus presentation, and the comparison is between times when a coping instruction was presented in response to a change in physiology (PT conditions) versus times when a coping instruction was presented in response to a pre-determined time schedule
(VT, FT; Other conditions). The expected outcomes of this question are detailed in Hypothesis 3. This question can be summarized as: does receiving and following a coping instruction have a greater effect when the instruction is physiologically-relevant, versus when it is physiologically-arbitrary?

Performance analysis of high stress conditions involved two indices representing performance: average time between consecutive deaths and average time between consecutive kills. Average time between consecutive deaths was calculated to represent performance on easily acquired and mastered skills. Evading death in this specific task is a skill that requires little training to obtain. Conversely, average time between consecutive kills obtained represents performance on more technical skills. Obtaining kills in this task is a skill that requires extensive prior experience and practice to master.

A post-hoc power analysis indicated medium power (0.634) at the level of $\alpha=0.05$. All results reported to be significant will be at this $\alpha=0.05$ level.

Results

Paired Student’s t-tests were conducted to compare the means of subjective ratings in all high stress conditions and subjective ratings in all low stress conditions. Analysis of STAI reports indicated that there was a significant difference between the means across high stress and low stress conditions. This supports that a higher state of stress was elicited in the Arms Race (high-stress) conditions compared to the Weapons Course (low-stress) conditions (Table 3.1). Comparisons of means for NASA-TLX also indicated that there was a significant difference between the means across high stress and low stress conditions, supporting a significantly higher
perceived cognitive workload elicited by the high-stress conditions compared to the low-stress conditions (Table 3.1).

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Mean</th>
<th>p-value</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STAI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Stress</td>
<td>21</td>
<td>1.75</td>
<td>p&lt;0.0001</td>
<td>0.4821</td>
</tr>
<tr>
<td>Low Stress</td>
<td>21</td>
<td>1.57</td>
<td></td>
<td>0.3997</td>
</tr>
<tr>
<td><strong>NASA-TLX</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Stress</td>
<td>21</td>
<td>50.3</td>
<td>p=0.0043</td>
<td>25.5598</td>
</tr>
<tr>
<td>Low Stress</td>
<td>21</td>
<td>41.83</td>
<td></td>
<td>21.8038</td>
</tr>
</tbody>
</table>

Table 3.1 Sample sizes, mean values, p-values, and standard deviations of participants’ overall state of stress and perceived cognitive workload.

In all following results, consistent abbreviations will be used to reflect conditions within physiological analysis. Specifically, “PT” will refer to conditions in which coping instructions were displayed according to substantial increases in physiological indicators of stress; PTH refers to those conditions which are performed under high stress (when participants play the Arms Race mode), and PTL refers to those conditions which are performed under low stress (when participants play the Weapons Course mode).

“Other” will refer to the remaining conditions, but its meaning changes slightly depending on the research question asked. When reporting data for times surrounding HR elevation, “Other” refers to conditions in which coping instructions were displayed either on a fixed time schedule, a variable time schedule, or not at all, and these conditions represent times in which HR was elevated but a coping instruction was not delivered. When reporting data for times surrounding instruction presentation, “Other” refers to conditions in which coping instructions were displayed either on a fixed or variable time schedule. These conditions represent times in which a coping instruction was delivered, but HR may or may not have been elevated. In both cases, OtherH refers to all of those conditions played under high stress (in Arms Race mode), while OtherL refers to all of those conditions played under low stress (in Weapons Course mode).
Repeated measures ANOVAs calculated to compare the differences between mean RMSSD values before and after moments of substantially elevated HR, using a 2 (stress level: high stress, low stress) x 2 (condition: Other, PT) design, reveal that there is no statistically significant difference (Figure 3.1, Table 3.2, Table 3.3).

![Graph showing Δ RMSSD surrounding peaks in HR](image)

**Figure 3.1** At times when HR was elevated, not receiving a message (Other) corresponds with a decrease in RMSSD—reflective of an increase in stress—following the peak in HR in the high stress condition, and no difference at all in the low stress condition. On the other hand, receiving a message at this time (PT) corresponds with an increase in RMSSD—reflective of a decrease in stress—following the peak in HR in both states of stress, but especially in the high stress conditions. However, these relationships are not statistically significant.

<table>
<thead>
<tr>
<th>RMSSD values surrounding peaks in HR</th>
<th>PTH</th>
<th>PTL</th>
<th>OtherH</th>
<th>OtherL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTH</td>
<td>0.7527</td>
<td>0.1826</td>
<td>0.5541</td>
<td></td>
</tr>
<tr>
<td>PTL</td>
<td>0.7527</td>
<td>0.3128</td>
<td>0.8034</td>
<td></td>
</tr>
<tr>
<td>OtherH</td>
<td>0.1826</td>
<td>0.3128</td>
<td>0.3907</td>
<td></td>
</tr>
<tr>
<td>OtherL</td>
<td>0.5541</td>
<td>0.8034</td>
<td>0.3907</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RMSSD values surrounding stimulus presentation</th>
<th>PTH</th>
<th>PTL</th>
<th>OtherH</th>
<th>OtherL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTH</td>
<td>0.8722</td>
<td>0.0608</td>
<td>0.1442</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3.2

<table>
<thead>
<tr>
<th></th>
<th>PTL</th>
<th>OtherH</th>
<th>OtherL</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSSD</td>
<td>0.8722</td>
<td>0.0608</td>
<td>0.1442</td>
</tr>
<tr>
<td>RMSSD</td>
<td>0.1014</td>
<td>0.1014</td>
<td>0.2047</td>
</tr>
<tr>
<td>RMSSD</td>
<td>0.2047</td>
<td>0.4938</td>
<td>0.4938</td>
</tr>
</tbody>
</table>

The mean RMSSD value representing the state of stress as a result of receiving a coping instruction at times of physiological stress and in a high-stress situation (PTH) is trending towards being significantly higher than the mean RMSSD of times when individuals were physiologically stressed and in a high-stress situation, but did not receive a coping instruction (OtherH). The mean RMSSD value representing the state of stress as a result of receiving a coping instruction at times of physiological stress and in a high-stress situation (PTH) is trending towards being significantly higher than the mean RMSSD of times when individuals were not necessarily physiologically stressed at the moment of instruction presentation, but were amidst an overall stressful situation (OtherH).

Analysis using repeated measures ANOVAs to compare the differences between RMSSD values before and after times of stimulus presentation, using a 2 (stress level: high stress, low stress) x 2 (condition: Other, PT) design, reveal that there is no statistically significant difference (Figure 3.2, Table 3.2, Table 3.3).

#### Figure 3.2

When a coping instruction is presented and followed at times not tied to elevations in HR (Other), there is a modest increase in RMSSD—reflective of a decrease in stress—when in a low-stress condition, and no difference at all when in a high-stress condition. When a coping instruction is presented and followed in response to elevations in HR (PT), there is a marked increase in RMSSD—reflective of a decrease in stress—in both states of stress, and even more so in the low stress conditions. However, these relationships are not statistically significant.
Repeated measures ANOVAs were conducted to analyze performance according to the number of instructions followed in a given condition (none, one or more, and four or more instructions followed). The least amount of time between kills occurred in the conditions where four or more instructions were both presented and followed, which indicates a higher frequency of kills and thus higher performance. The difference between the average times between kills for all three categories, however, was not statistically significant (Figure 3.3, Table 3.4, Table 3.5).

Table 3.3 Sample sizes, mean values, and standard deviations of participants’ RMSSD values surrounding peaks in HR and stimulus presentation

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Mean (ms)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Δ RMSSD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>surrounding peaks in HR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OtherH</td>
<td>6</td>
<td>-55.72</td>
<td>186.743</td>
</tr>
<tr>
<td>OtherL</td>
<td>6</td>
<td>1.03</td>
<td>59.3347</td>
</tr>
<tr>
<td>PTH</td>
<td>4</td>
<td>44.49</td>
<td>19.5551</td>
</tr>
<tr>
<td>PTL</td>
<td>4</td>
<td>19.23</td>
<td>42.8380</td>
</tr>
<tr>
<td><strong>Δ RMSSD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>surrounding stimulus presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OtherH</td>
<td>12</td>
<td>2.11</td>
<td>49.7429</td>
</tr>
<tr>
<td>PTH</td>
<td>21</td>
<td>24.87</td>
<td>61.1735</td>
</tr>
<tr>
<td>PTL</td>
<td>11</td>
<td>75.29</td>
<td>92.6773</td>
</tr>
<tr>
<td>PTL</td>
<td>6</td>
<td>86.29</td>
<td>215.6409</td>
</tr>
</tbody>
</table>

Figure 3.3 Average time between kills was lowest, representing the highest frequency of kills and highest performance by this metric, when four or more instructions were followed. The difference between this
mean and the mean when zero instructions were followed is indistinguishable, however, and did not come close to statistical significance.

<table>
<thead>
<tr>
<th>Time between kills (s)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1+</td>
<td>4+</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0.2434</td>
<td>0.8042</td>
</tr>
<tr>
<td>1+</td>
<td>0.2434</td>
<td></td>
<td>0.3764</td>
</tr>
<tr>
<td>4+</td>
<td>0.8042</td>
<td>0.3764</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time between deaths (s)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1+</td>
<td>4+</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0.0185*</td>
<td>0.3005</td>
</tr>
<tr>
<td>1+</td>
<td>0.0185*</td>
<td></td>
<td>0.7138</td>
</tr>
<tr>
<td>4+</td>
<td>0.3005</td>
<td>0.7138</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.4** The mean amount of time (in seconds) between kills does not differ significantly when comparing the mean values between conditions representing various levels of intervention adherence. The mean amount of time (in seconds) between deaths is significantly higher when one or more of the coping instructions within the intervention are adhered to (1+) compared to when none of the coping instructions are adhered to (0). This is the only significant difference among means as a result of a post-hoc comparison of means.

Repeated measures ANOVAs for time between deaths revealed more time between deaths, on average, when any number of instructions were presented and followed, compared to when zero instructions were presented and/or followed. More time between deaths represents a lower frequency a deaths, and thus higher performance. These results again did not reach statistical significance, but were trending towards a significantly higher amount of time between deaths when one or more instructions were followed compared to when zero instructions were followed (Figure 3.4, Table 3.4, Table 3.5).
Figure 3.4 Average time between deaths was highest, representing the lowest frequency of deaths and highest performance by this metric, when one or more instructions were followed. The difference between means when one or more instructions were followed and when zero instructions were followed is not statistically significant, but is approaching $\alpha=0.05$

<table>
<thead>
<tr>
<th>No. of instructions followed</th>
<th>N</th>
<th>Mean (sec)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time between kills across conditions</td>
<td>0</td>
<td>38</td>
<td>130.70</td>
</tr>
<tr>
<td></td>
<td>1+</td>
<td>29</td>
<td>171.29</td>
</tr>
<tr>
<td></td>
<td>4+</td>
<td>6</td>
<td>115.39</td>
</tr>
<tr>
<td>Time between deaths across conditions</td>
<td>0</td>
<td>45</td>
<td>19.76</td>
</tr>
<tr>
<td></td>
<td>1+</td>
<td>34</td>
<td>22.94</td>
</tr>
<tr>
<td></td>
<td>4+</td>
<td>8</td>
<td>22.09</td>
</tr>
</tbody>
</table>

Table 3.5 Sample sizes, mean values, and standard deviations of participants’ performance metrics, including time between kills and time between deaths across conditions

In summary, physiological results broadly suggest that receiving and following externalized assistive coping strategies at times of elevated HR produced better downstream physiological recovery than both not receiving a message at times of elevated HR and receiving a message at a time not based in physiological variables. Although these values did not reach the threshold of statistical significance, the relationships reported align with expectations.
Performance results suggest overall that greater intervention adherence contributes to more optimal performance on the acutely stressful task evaluated.

Discussion

The purpose of the reported study was to provide evidence to support the inclusion of explicit, externalized coping strategies during the completion of an acutely stressful task in order to improve stress recovery as indicated by near-real time physiological indicators and subjective perceptions. I was also interested in the influence of coping instructions on task performance. The primary motivation of asking these research questions and pursuing the experiment reported was the lack of relevant evidence supporting the inclusion of coping instructions. Despite this dearth of literature, there is a growing focus on coping strategies employed in high-stress settings to manage stress (in healthcare in particular) (Anton et al., 2015; Arora et al., 2009), and cognitive theories supporting the externalization of strategy implementation (Schnotz & Kürschner, 2007). Coping strategies employed by surgeons specifically have reportedly included cognitive strategies and stopping, stepping back, and taking a deep breath most commonly (Arora et al., 2009). The concept of stopping and taking a deep breath has appeared both in interviews with surgeons and an open-ended poll posted for this project inviting responses directly from CSGO players (unpublished).

The current work utilized a systematic and controlled environment, and evaluated a population of healthy undergraduate students as they completed a computer-based task. The cognitive processes recruited through the acutely stressful task used deliberately approximate many underlying cognitive processes relied upon in high-risk and highly stressful professions. High visual-spatial demands, dynamic decision-making, multi-tasking, and intentional and active
attention allocation are examples of some of the demands represented in this task and in the domain-specific tasks in an industry such as healthcare, and more specifically, surgery. By eliciting stress in a highly controlled setting that recruits cognitive demands similar to those experienced by healthcare practitioners, we can be confident of the minimal influence of extraneous load incurred by the environmental surroundings, and begin to glean insights and extend implications to the healthcare setting.

Although I did not recruit practicing healthcare practitioners, I believe the CLT theory utilized in this approach can appropriately represent cognitions of participants under acute stress, while also appropriately projecting expected cognitions of future intended participants, individuals in healthcare. I am confident in this due to the prior utilization of the CLT framework to a variety of applications, including medical education, e-learning, clinical diagnosis, and ergonomics settings (Galy & Melan, 2015; Van Merriënboer & Sweller, 2010).

As the results of the physiological data analysis show, the differences between means was not statistically significant for any of the pairs within physiological data analyzed surrounding peaks in HR or stimulus presentation. However, results of the standard deviations within in each condition for data surrounding HR show that the PTH condition in particular had a substantially higher mean RMSSD value (representative of a lower state of stress) and a substantially lower standard deviation value compared to the other conditions analyzed. One interpretation of these observations could be that in a high stress situation, receiving a coping instruction at the height of physiological stress (PTH condition) most consistently produces a beneficial physiological outcome. Our results suggests that the increased benefit associated with receiving coping instructions at the peak of physiological stress (PTH condition) results in a different and more
positive, but not statistically significant, outcome than not receiving a coping instruction at this time (OtherH conditions).

Physiological data surrounding times of stimulus presentation reveal that the OtherH condition has the lowest mean and lowest variability. This finding could suggest that in a high-stress situation, receiving a coping instruction at a time when you are not in fact physiologically stressed (OtherH conditions) could more consistently lead to less positive effects compared to the other conditions.

Results of the t-tests for data surrounding stimulus presentation also provide additional insight. These data suggest that the difference between mean RMSSD values among the PTH and OtherH conditions is closely approaching significance at the $\alpha = 0.05$ level, which supports that there is an increased benefit associated with receiving coping instructions in stressful situations when an individual is indeed physiologically stress (PTH condition), compared to times when they may not truly be physiologically stressed (OtherH conditions). Also approaching significance, but less substantially, are the comparisons of means between PTH and OtherL conditions, and the comparisons of means between the PTL and OtherH conditions, suggesting that receiving a coping instruction when physiologically stressed (PTL and PTH conditions) is likely more valuable and beneficial than receiving one when one is not physiologically stressed (OtherL and OtherH conditions), regardless of whether the overall situations represents one of low or high stress.

The results of physiological data analysis here are not statistically significant and our corresponding hypotheses are thus not supported. However, I believe the trends in my findings merit further exploration, particularly given the paucity of literature in this area, and given the potential implications for high-risk work. Our data appear to trend towards the idea that germane
load can in fact be reduced through the appropriate administration of explicit and externalized coping strategies, but further research with larger sample sizes must be completed to make more assertive claims in this regard.

In my performance analysis, the index representing automatic skill acquisition appeared to improve with greater adherence to the externalized coping instructions, contributing some support to my expectation that reducing germane load by minimizing the participants’ requirement to arrive at coping strategies on their own can improve automatic or less effortful skill. This finding, although not reaching significance, lends support to Sweller’s original definition of germane load (Sweller, 1988).

While the exact task presented in this study and performance measures analyzed would not be performed in the healthcare setting, we can begin to appreciate the representative cognitive processes recruited through the task, along with the tightly controlled environment and the potential implication for the performance indices selected. For example, the lack of association between adherence with this intervention and performance on a technically-intensive skill could generalize to suggest that novel or newly acquired skills, broadly speaking, may not benefit from deliberate efforts to reduce stress in real time. If we consider the automatic and controlled processing theory and how it relates to optimizing human performance under stress (Fisk, Ackerman, & Schneider, 1987), newly-acquired and difficult-to-master skills may thus require an individual’s full attention and greater cognitive resources in general. There is greater potential that stress reduction efforts via adhering to externalized coping strategies will be successful for those skills which are already mastered and more automatically executed in nature, requiring less cognitive resources overall.
Although this study recruited a convenience sample, we can infer potential contributions and direct benefits this research may generate for the healthcare domain. In particular, the knowledge gained from this study will contribute to the foundation of future work. Results will influence impending evaluations of related concepts among more specialized populations, including medical students, residents, fellows, and attending healthcare practitioners. Exploring the basic relationships among the physiological markers of acute stress and the associated cognitive functioning and performance among a homogenous convenience sample has allowed greater control at this early stage of investigation.

In addition to the long-term benefits this line of research may have on healthcare training and practice, our methods represent direct benefits to healthcare informatics in terms of data generation. Physiological sensors passively and continuously generate up to thousands of data points every second. The generation of such a large bank of data points could feasibly be used to inform neural networks and machine learning algorithms to ensure a reliable and accurate method of real-time, automatic stress detection. Advances such as this could benefit the field of healthcare at large by creating a comprehensive and objective profile of patients’ health states as they fluctuate in real time, for example. Passive sensors’ constant data generation could also be utilized in a similar way to create a profile of acute stress and predict patterns of cognitive and performance changes among healthcare practitioners. Investigation into this vision is already underway through projects such as ACLAMATE (Pappada et al., 2016), which leverages psychophysiological data acquired through non-invasive sensors to generate an objective, temporally-specific objective measure of stress, cognitive workload, and performance among medical staff.
Medical staff at our own teaching hospital have confirmed the potential utility of the larger question that this study has begun to answer regarding an acute stress management intervention, if applied to medical residents and fellows during their medical training. A brief contextual inquiry and subsequent contextual analysis has generated qualitative data to support the inclusion of biofeedback approaches to stress recognition and management, as well as specific coping mechanisms attending medical staff tend to rely on and view to be most appropriate and effective. Three of the four surgeons (75%) who explicitly provided ideas for the form of feedback in this theoretical intervention specifically mentioned measuring HR, and three of them (75%) also mentioned displaying biofeedback in a simple visual interface of trending information. Notably, five of the six attending surgeons interviewed (83%) specifically mentioned the potential utility in building this type of intervention into trainings for residents and fellows. Focusing on breathing as a means of coping efficiently was a recurring theme that arose from these interviews as well.

Despite the potential application of this work, there are a number of important limitations. The task selected, CSGO, limited the options for analysis, in that it was difficult to derive an index that properly represented performance. For example, not only is obtaining kills a technical skill that is extremely difficult to master (even after playing 4 of the 10-minute high stress conditions for a total of 40 minutes), but also performing well in this sense amplifies the potential to perform well in the future. As the player obtains more kills, they move up in level. Reaching higher levels rewards the player with better weapons, making it easier to continue to get more kills. In this way, performing well according to the index of number of kills makes it easier to continue performing well or better. For both of these reasons (technically-intensive, self-perpetuating advancement), the time between kills or its inverse, frequency of kills, may not
the best index of performance on this task. Further, 19 of 21 participants had never played before, making it extremely difficult to acquire skills such as appropriate weapon mechanics. Even so, 18 of the 21 participants (86%) were able to get at least one kill in at least one condition, with 67 of the 84 conditions (80%) reporting at least one kill. The average number of kills per condition was 4.88 across participants, so this skill was demonstrably possible to acquire to some degree, even by individuals who had no experience with this task.

Performance analysis of low stress conditions also had limitations. Each low stress condition for each participant followed the same format, in a pre-determined and scripted sequence of required actions to complete in succession. Simply quantifying how far an individual advanced through the tutorial in each condition is thus not an appropriate representation of performance. Every participant advanced further each time they experienced the low stress condition, showing a learning effect, despite the initial training on how to play the game.

The primary limitation associated with analysis of physiological changes and the lack of significant findings within these measures was the low sample size. Even though I had a considerable sample size overall of 21 participants each completing all eight conditions, filtering the data in an effort to select physiological events that were most temporally isolated and minimally confounded resulted in very few individual data points to compare, which likely contributed to the trending, but not statistically significant, nature of our results. The large variance observed in the statistical reports poses an additional limitation, which may have also resulted from small sample sizes and contributed to a lack of statistically significant results.
Conclusions

In conclusion, this study sought to address whether we could improve physiological and cognitive signatures of acute stress, as well as task performance, by delivering coping strategies during the completion of an acutely stressful task. The primary motivation of this endeavor was the lack of relevant evidence supporting the inclusion of coping instructions. The current study evaluated a population of healthy undergraduate students in an experimental setting as they completed a computer-based task. The nonsignificant trends observed in the data suggest that receiving coping strategies at times of elevated HR produced better downstream physiological recovery than both not receiving a message at times of elevated HR and receiving a message at a time not based in physiological variables. Performance results suggest overall that greater intervention adherence contributes to more optimal performance on the acutely stressful task evaluated.

The use of coping instructions as a primary acute stress intervention alone is not supported by the findings of this study, but the potential for its benefit in combination with highly successful interventions already in use can be appreciated. Our findings have direct implications for the presentation of assistive information in acute stress settings. The acquisition of these results in a tightly controlled setting lays the groundwork for a similar evaluation within a more applied setting, such as surgical simulation. This foundational work is critical, as an inappropriate design of coping instructions could lead to too many interruptions, distracting or extraneous information, and ambiguous or unhelpful coping instructions (Kontogiannis & Kossiavelou, 1999), pointing to the potential to exacerbate stress rather than ameliorating it (MacLean, Roseway, & Czerwinski, 2013).
Future work will evaluate the benefit of a biofeedback-based intervention for the purposes of stress reduction and performance enhancement among surgical residents and fellows. Crucially, future work will also evaluate the benefit of the combination of biofeedback and a series of supplementary coping instructions to determine whether the combination of approaches could have benefits above and beyond either intervention administered alone.

Before getting to the point of systematic evaluation of intervention components, I wanted to ensure that the design and structure of the concept was well-informed. To gain the most insight into appropriate design considerations for a cognitive aid intended for surgeons, I needed to learn the thoughts and perspectives of the subject matter experts themselves. Another primary concern was ensuring that the task completed induced enough stress in order to measure any degree of stress reduction incurred by the intervention. Thus the next phases of my dissertation work involved interviewing attending surgeons and validating a stress-inducing task.
Chapter 4a
Contextual Inquiry and Analysis

Abstract

Before initiating the systematic evaluation of biofeedback components among a demographic of healthcare practitioners, the design must be appropriate. The purpose of the contextual inquiry and analysis phase of this dissertation work was to better define user requirements and develop a use case for biofeedback in near-real time for applications during surgery. This study was conducted to ensure that the results of my larger studies would be applicable and useful during an operation.

Introduction

As is the case with any device designed for cognitive support, beyond establishing usability, it must further address a need for its use among the target population. If the end users have no interest in using such a device or interface, ensuring optimal functionality and usability is futile. In the healthcare industry in particular, operating rooms are overrun with displays competing for the attention and cognitive resources of healthcare practitioners, and we cannot afford for an additional display to be distracting or obtrusive. Despite the historical view that acknowledging stress reflects a weakness or incompetence on the part of the surgeon, recent investigations have shown that healthcare centers are in fact interested in a stress management training program (Anton et al., 2015).

While the need for stress management may be increasing, it is still unclear whether an intervention such as biofeedback would be welcomed in the context of surgery. In light of this, this phase of my dissertation involved interviewing the end users themselves to get a better sense
of the potential utility and need of a visual biofeedback interface during surgery. It is critical to
take into consideration the context in which a device or display will be integrated, and the
characteristics of the users interacting with it. The context and the users are key components of
determining a product’s overall usability, which is typically determined by the product’s features
and combination of features, its functionality, and its usefulness. These concerns are especially
important in the high-risk socio-technical system of healthcare (Corrao, Robinson, Swiernik, &
Naeim, 2018).

The questions comprising the contextual inquiry process were intended to determine
whether such an interface would be appropriate and incorporated by the end users, and to
identify parameters, and features of those parameters that will facilitate use, to optimize the
design. Learning about perceived barriers was also an important goal of the interviews, in order
to determine pitfalls to avoid. In order to arrive at this information, I adopted a structure akin to
the critical incident technique (Flanagan, 1954). The goal of critical incident technique
approaches is to collect and analyze reports of actual behaviors—in this case, actual times when
attending surgeons have experienced acute stress—and extrapolate from those reports to interpret
the role certain behaviors have on job performance (Anderson & Wilson, 1997).

Materials and Methods

Participants included seven surgeons practicing at Carilion Roanoke Memorial Hospital
(CRMH). All seven participants were male, and data collection concluded after saturation of
responses was detected (Ando, Cousins, & Young, 2014; G., Strauss, & Strutzal, 1968). The
determination of reaching saturation can be loosely identified as exhausting observable themes in
the data. Transcribing and determining themes in the data concurrently with data collection
allowed for the opportunity to observe emergent themes and monitor to detect the cessation of newly emerging themes (Ando et al., 2014). Of the seven interviewees, six were attending surgeons with specialties representing surgical oncology, ophthalmology, and colorectal, pediatric, orthopedic, and general bariatric and hepatobiliary surgery. One participant was a resident, in post-graduate year 3, specializing in plastic and reconstructive surgery.

Face-to-face, semi-structured interviews took place in the surgical lounge of CRMH in July, 2016 and ranged from 4:56 to 17:12 in length, while spanning 9:48 on average. All interviews were recorded with handheld recording devices and manually transcribed. The interviews were approved by and conducted in accordance with the guidelines of the Carilion Clinic Institutional Review Board, and subjects were aware that by answering interview questions, they were providing verbal consent to participate.

Interview questions were designed to be open-ended, with pre-determined follow-up questions in place to ensure complete answers (Table 4a.1). The contextual inquiry and analysis approach was adopted in order to extrapolate how the intended users of the biofeedback interface do, need, and think. This process involves paying attention to the role of emotion, and the overall context in the design of a new device or interface. Contextual analysis of responses from interviewees that followed was qualitative in nature. Analysis involved first transcribing interviews, and then categorizing responses within the context of the work domain and the individuals interviewed. Categories were not pre-defined, allowing for the observation and generation of emergent categories of responses. As interviews were collected and transcribed, I grouped responses according to these emergent categories, and continued conducting interviews until it became evident that no additional themes were emerging, indicating saturation in the data (Ando et al., 2014).
Are there ever times during surgeries when you feel like things are outside of your control? Times when, even if you had all the resources and time in the world, the nature of the situation is overwhelming?

How often would you say you feel this way?

Would it be fair to call such a situation “stressful”?

Can you share a specific story that illustrates this? Regardless of whether the outcome was positive or negative.

How do you deal with this and/or restore control?

Thinking back on the story you shared, during time-sensitive and serious (stressful) procedures, what sorts of things do you pay the most attention to?

Where do you allocate visual attention? What do you listen for the most in the environment? What types of haptic/touch sensations do you attend to the most?

Are you aware of any of your own physiological changes that take place that you’re aware of at times like this?

How do you address these physiological changes, if at all?

Can you think of a way this could be externally represented that might make this process easier/more reliable for you?

Imagine a system that could convey biometric information.

Can you describe what this might look like to you?

What information would be helpful to you?

How would you imagine it should be presented?

When would this information be most helpful?

What might some foreseeable barriers to this system be from your perspective?

Table 4a.1 Interview questions and follow-up questions comprising the semi-structured interviews. Not all questions were asked, and they were not necessarily addressed in the same order from interview to interview, depending on how the conversation unfolded.

Results

Data collection was stopped after saturation of responses became noticeable, after interviewing seven total participants. None of the following results are reported in terms of statistical significance.

In response to the first question and its corresponding follow-up questions concerning the acknowledgment and frequency of stress, answers were inconsistent. While five of the seven
interviewees replied that they do experience stress in the operating room during surgeries, one of those individuals estimated he experienced stress about 20% of the time, while the others mentioned “infrequently” or “very rarely.” Interestingly, two of the seven interviewees replied that they never experience stress while operating. These participants were asked to think back to their training years when they may have felt stress more often, or imagine the way their residents feel today, and respond to the remaining questions with those feelings in mind.

When asked to share a story about this, many individuals made reference to the source of this stress. Responses were categorized primarily as a lack of control or unanticipated events precipitating most of the stress these surgeons experience (N=4). Other sources mentioned included poor communication, lack of experience of the team, preparing to operate, and problems with instruments and equipment (N=1). In an effort to regain control, participants mentioned deep breathing (N=3), taking a step back (N=2), slowing down and planning (N=2), and regulating emotions (N=2). When prompted about what they pay attention to at moments like this, when under extreme stress, all respondents providing an answer said they focus on the patient (N=3).

Participants were then asked about their awareness of their own physiological changes during stressful periods, two participants indicated they do not focus on their own physiological response, while one participant mentioned simply that he becomes aware of these changes after the fact. The remaining interviewees all referenced their awareness of increased heart rate (N=4). Other responses included changes in breathing (N=2), an anxious feeling (N=2), and muscle tension and headache (N=1).

When interviewees were asked whether externally representing their own physiological changes could be helpful to them, answers varied. One participant indicated that this would be
helpful to him, two suggested it could be “if something could be done about it”, and the remaining two participants offered that while it would not be helpful for them, it could be helpful for trainees. The remaining follow-up questions were also open-ended and produced a range of ideas for what a helpful biofeedback system might look like, general and specific concerns about this idea, and additional comments.

Interviewees primarily indicated a simple display (N=2) conveying heart rate information (N=4) with embedded coping indicators (N=2). Modalities suggested included visual, auditory, and haptic, while one participant specifically suggested avoiding auditory inputs. Despite these suggested, overall concerns were that this type of intervention would be distracting (N=3) and potentially more stressful (N=2) if not designed properly. Another concern with modality raised by one interviewee was the potential public nature of the surgeon’s physiological information. One surgeon suggested that displaying or announcing the physiological status of the operating surgeon could be suggestive of their stress state as well. Their concern was that the team’s knowledge of the surgeon’s stress could be embarrassing for the surgeon and detrimental for the team at large.

In general, there was consensus among the six attending surgeons interviewed that a biofeedback interface with supplemental training on how to cope could be beneficial for trainees in the early learning stages (N=6). This concept arose naturally across all interviews, and was never intentionally built in to the set of questions used as a guide.

Discussion

Ensuing qualitative contextual analysis of the seven participating surgeons’ responses to my face-to-face interview questions helped to shape the direction this dissertation work took and
informed both my mental model and the design of the following studies. The suggestion by two participants that biofeedback would be helpful “if something could be done about it” called attention to the key factor of perception of control, and its role in stress. Uncontrollability plays a major role in stress induction among this population, and it became another area to include in the development of my stress-inducing task (see Chapter 4b). This concept also contributed to my focus on external sources of control as a means of stress reduction, reinforcing my interest in including a directive coping supplement. Specifically, it became clear that a biofeedback intervention with the additional component of integrated coping instructions could be useful, but the interviewees who I imagined to be end users made it clear that they were not in fact the appropriate target users. The attending surgeons I interviewed almost unanimously supported the use of biofeedback as a stress management approach for surgical trainees in the midst of residencies and fellowships. The most compelling aspect of the results was its natural convergence. Although I had not anticipated experience or expertise being a factor in these conversations, each attending surgeon independently concluded that the intervention discussed would be helpful for trainees.

Interviewees also brought attention to certain barriers of such an intervention that I had not previously considered, including the importance of keeping physiological information confidential and limiting others’ knowledge of the user’s stress levels by keeping the feedback private. This concern, along with the concern of cognitive overload and the potential for distraction, raised my interest in introducing at least one component of the intervention through an alternative modality. While haptic, visual, and auditory modalities were all contenders, equipment and design limitations determined that the coping instructions would be auditory in nature. If biofeedback were to be administered by the visual modality, interviewees specifically
supported using heart rate and/or respiratory feedback, displayed in a format compatible to other visual inputs, such as patient vital signs. With this in mind, my future studies involved using trending heart rate as the primary visual display of biofeedback information.
Chapter 4b
Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task

Abstract

The purpose of this project was to adapt an existing computer-based task called Multi-Attribute Task Battery (MATB-II), developed by NASA and frequently used to induce acute stress among air craft crew members and general populations, for use in medical populations. I gathered continuous electrocardiography (ECG) data while medical students completed four different versions of the MATB-II of varying difficulties alongside questions probing medical knowledge, comprising a new task called medically-focused multitasking game (MFMG). After completing each version, participants responded to questionnaires to assess subjective states of stress (State Trait Anxiety Inventory for Adults) and cognitive workload (NASA Task Load Index). Responses to these questionnaires, physiological data from continuous ECG, and overall performance scores were combined to determine one version of MFMG that represented the highest level of elicited stress, and one that represented the lowest level of elicited stress. The results of this pilot study are promising, and have converged to reveal one high-stress and one low-stress version of MFMG, which will later be used to induce acute stress in biofeedback intervention studies among surgical residents and fellows. Beyond this specific application, MFMG can have broader applications in measuring acute stress induction and/or reduction among populations of healthcare practitioners.
Introduction

Acute stress is a well-documented problem for surgeons and surgical performance, especially novice surgeons (Arora, Sevdalis, Nestel, et al., 2010; Hassan et al., 2006; Healy & Tyrrell, 2011). Excessive levels of intraoperative stress are associated with impaired cognition (Wetzel et al., 2006) and psychomotor performance (Arora, Sevdalis, Aggarwal, et al., 2010). Surgery represents an acute care setting particularly susceptible to errors (Gruen et al., 2006), which may be exacerbated by excess stress. For these reasons, managing acute stress is imperative to optimize surgical performance and limit medical errors.

Biofeedback is one example of a quickly advancing technological tool that can mitigate adverse effects of acute stress (Kennedy & Parker, 2018a). Biofeedback training, often conveying the user’s heart rate in real time, enables the user to manage their physiological activity to improve health and performance (Ortiz-Vigon Uriarte et al., 2015). As information is fed back to the user, voluntary and involuntary alterations in cognition, emotion, and behavior affect physiology, and the process continues iteratively (Schwartz, 2010). By monitoring physiological indicators of stress in real time, users could potentially learn to preemptively acknowledge and manage elevating stress levels.

Effective use of visual and auditory biofeedback among soldiers in simulation (Bouchard et al., 2012) offers convincing evidence for a similar benefit during acute stress simulation-based training for novice surgeons. Among surgical populations, biofeedback has also been shown to effectively reduce chronic stress levels (Lemaire, Wallace, Lewin, de Grood, & Schaefer, 2011). What remains to be seen is whether acute stress management can be accomplished using biofeedback in surgery.
Answering this question in the real-world setting has potential to introduce too much variability and confounding factors, minimizing the degree of control over the situation. The gap outlined represents one that first must be investigated in a setting free of excessive external influences, relying on a lab-based design. To reach these aims, a computer-based task must be investigated.

Since biofeedback in this capacity functions as an acute stress management tool, to evaluate its effectiveness we must first induce the stress we aim to mitigate. A variety of stress-inducing tasks evaluated through a meta-analysis of 208 lab-based experiments revealed that those containing uncontrollable and social-evaluative elements were associated with the largest increases in physiological stress (Dickerson & Kemeny, 2004). Given the uniqueness and complexity of the specific target population and its corresponding work domain, we must consider how researchers in the past have induced acute stress in other high-risk professions with stress levels approximating those seen in surgery.

Multi Attribute Task Battery (MATB-II), a well-validated and reliable tool, contains the key element of uncontrollability and was originally designed by NASA to induce stress among aircraft crew members and pilots (Comstock & Arnegard, 1992). Its applicability to other complex work domains with frequent exposures to acute stress is a key feature of MATB-II, due to its focus on simultaneous monitoring of multiple tasks at once. However, MATB-II is not specifically medically-relevant and does not address the crucial cognitive process of memory recall critical to medical decision-making.

To address these gaps, our lab adapted the MATB-II into a new task incorporating medical knowledge. This new tool, referred to as Medically-Focused Multitasking Game (MFMG), integrates three of the four MATB-II subtasks with timed medical questions. By
combining MATB-II subtasks with medically-relevant questions, I have developed a new task that requires multitasking, ongoing monitoring and adjustment, and medical knowledge that is appropriate for the intended setting and population: surgical residents and fellows.

Prior to using MFMG in experimental settings, it is necessary to assess its capacity to appropriately induce stress, and to determine the impact of the medical questions on stress performance. MATB-II output represents performance on multiple subtasks, each of which can be manipulated to arrive at various versions of task difficulty levels. The degree of difficulty of the overall MFMG, which is largely determined by the degree of difficulty of MATB-II, has the potential to contribute to a variety of response patterns, some of which are undesired. Generating a level of difficulty that is too under- or too overwhelming could easily lead to boredom and/or disengagement, respectively. The task, therefore, needs to be stressful enough to induce the appropriate physiological and cognitive response, but not so stressful that disengagement ensues.

This study aims to address this process of realizing two independent versions of MFMG that represent appropriately low and high levels of stress in order to evaluate a stress management intervention in a population of surgical residents and fellows. I pilot tested versions of MATB-II in combination with medical questions (together comprising MFMG) among a group of medical students interested in the surgical domain. Analysis of physiological indicators of stress, subjective responses, and overall performance score determined the appropriate amount of stress for one low stress and one high stress version of MFMG.

I expect that subjective, objective, and performance results will align and thereby indicate which condition is considered the least stressful and which is considered the most stressful. This knowledge will inform the experimental design of future studies evaluating the most effective timing of presentation of components of an integrated biofeedback display with supplemental
coping instructions. The display evaluated in future studies will have long-term implications for the appropriate integration of biofeedback as an acute stress management tool for novice surgeons during simulated stressful scenarios.

Materials and Methods

Participants

A total of 10 medical students (5 females) were recruited to participate in this study. The participants included students across all years (3 first-years, 2 second-years, 2 third-years, 3 fourth-years), and 9 of the 10 participants were right-handed. The experiment was approved by and conducted in accordance with the guidelines of the Carilion Clinic Institutional Review Board, and subjects provided verbal consent before the study began.

Materials

Electrocardiography (ECG) signals were recorded using a 3-lead Mobile Cardio acquisition system (MindWare Technologies LTD, Gahanna, OH). This system acquired continuous heart rate (HR) and heart rate variability (HRV) data by affixing three lead-shielded electrode leads to disposable ECG electrodes transmitting voltage signals from the chest. Data was transferred via SD card to MindWare’s BioLab software and HRV application for analysis.

The MATB-II task was completed on a Dell desktop computer, while the medical questions were displayed and attended to simultaneously on an 8-inch Asus ZenPad 8.0.
Procedure

After going through the information sheet and providing verbal consent, participants were equipped with ECG electrodes and five minutes of baseline data were collected as the participant relaxed. Participants then viewed a 12-minute tutorial overviewing the task requirements of the following MATB-II subtasks: resource management, system monitoring, and tracking (Figure 4b.1).

![Figure 4b.1 MATB-II interface. The MFMG task incorporated the following subtasks seen in this interface: system monitoring (top left), tracking (top center), and resource management (bottom center). The communications subtask was not included (bottom left). Medical questions were displayed on a separate screen.](image)

The main goal of the resource management task is to maintain 2500 units worth of fuel in the main tanks (A and B), despite their constantly depleting nature and occasionally broken pumps. The system monitoring task requires participants to detect and correct when the gauges deflect too far from the center, and when the lights are inappropriately on or off. To successfully perform the compensatory tracking task, participants must use a joystick to maintain the aircraft’s position (indicated by the circle) within the dotted square in the center. After summarizing the tasks and asking any questions, participants spent 20 minutes practicing the entire task, including all three MATB-II subtasks and the timed questions on the Asus tablet. Questions in this familiarization phase were not medically-oriented, but were comprised of random trivia knowledge with the purpose of acquainting participants with the appropriate
gestures and multi-tasking requirements. Participants then filled out a brief demographic questionnaire with no identifying information and form Y-2 of the State Trait Anxiety Inventory for Adults (STAI) to assess the stable trait of overall stress levels.

The experimental phase consisted of completing four, 5-minute versions of the MFMG task. See Table 4b.1 for details on the frequency of events contributing to the level of demand. Medical questions represented a variety of specialties, ranging from neurosurgery to gerontology to psychiatry, for example. One medical student and one school of medicine faculty member reviewed the questions and determined them to be appropriate for gauging knowledge of medical residents and fellows, with potential difficulty foreseen in assessing medical students’ medical knowledge. Following each condition, participants responded to subjective measures of state stress (State Trait Anxiety Inventory for Adults, form Y-1; STAI) and cognitive/mental workload (NASA Task Load Index; NASA-TLX). At the completion of the fourth condition, participants weighed each workload dimension in the NASA-TLX.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Tracking</th>
<th>System Monitoring</th>
<th>Resource Management</th>
<th>Medical Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATB-A</td>
<td>Medium default</td>
<td>30 deflections/min</td>
<td>1-2 pumps fail for 15 seconds every minute</td>
<td>Randomized order</td>
</tr>
<tr>
<td>MATB-B</td>
<td>Medium default</td>
<td>20 deflections/min</td>
<td>1-2 pumps fail for 5 seconds every minute</td>
<td>Randomized order</td>
</tr>
<tr>
<td>MATB-C</td>
<td>Easy default</td>
<td>10 deflections/min</td>
<td>1 pump fails for 15 seconds every minute</td>
<td>Randomized order</td>
</tr>
<tr>
<td>MATB-D</td>
<td>Easy default</td>
<td>2 deflections/min</td>
<td>1 pump fails for 5 seconds every minute</td>
<td>Randomized order</td>
</tr>
</tbody>
</table>

Table 4b.1 Characteristics for each subtask of MFMG. The frequency of events within each MATB-II subtask differs depending on the version of MFMG being played, while the medical assessment questions are always randomized. The medical assessment portion of MFMG should not differ in difficulty or affect the overall difficulty across versions.

Analysis

Total scores for the STAI and NASA-TLX were calculated using Excel and the means were compared using repeated measures ANOVAs calculated in JMP Pro 12 by SAS Institute,
Inc. (Cary, NC). These data were used to provide a more subjective indicator of the most and least difficult versions of the overall task.

HR and HRV analysis were conducted using MindWare’s BioLab acquisition and HRV analysis modules. Calculations of time-domain features of HRV such as standard deviation from normal-to-normal (SDNN), root mean square of the successive differences (RMSSD), and percentage of consecutive N-N peaks differing by 50 milliseconds (pNN50) provide us with measures of overall variability and beat-to-beat variability. Five minutes of data were selected from the 20-minute practice period, from the time range of 10-15 minutes, to derive vanilla baseline values of each of the metrics analyzed. SDNN, RMSSD, and pNN50 values were then normalized for every individual based on their values during the vanilla baseline period, and aggregated across participants. Results from repeated measures ANOVAs using JMP Pro 12 provide a more objective ranking of stress.

Performance analysis using Microsoft Excel and JMP Pro 12 provides an overall performance score by assigning equal weight to the three MATB-II subtasks and combining them. Performance on the resource management subtask was represented by calculating the average deviation from 2500 units in tanks A and B. To represent performance on the system monitoring task, I calculated the proportion of correct responses to deflections. Performance on the tracking task was represented by the root-mean-square (RMS) error. Finally, performance on the medical assessment questions was evaluated by the proportion of correct responses. Following these analyses, performance on each subtask was normalized according to overall performance on that individual’s subtask in the practice period, creating a z-score, with the exception of the medical assessment subtask. This exception was made because questions administered during the practice period were intentionally non-medical and represented random
knowledge. The nature of the practice questions was not comparable to that of the medical questions, or representative of the pre-acquired knowledge this subject pool was expected to have. The z-scores from the MATB-II subtasks and a raw score representative of medical assessment performance were then summed to create one value representing overall performance, with each subtask receiving equal weight. All performance scores were averaged across participants within each version of MATB-II, and those means were compared using repeated measures ANOVAs in JMP Pro 12.

Results

Subjective Indicators

Subjective states of stress and cognitive workload were analyzed by conducting repeated measures ANOVAs on scores on the STAI and NASA-TLX, respectively. While results from the STAI show no significant differences between conditions, the results from the NASA-TLX do reveal significant differences (Figure 4b.2). Specifically, comparisons of means using paired t-tests indicated that there was a significant difference between the mean perceived cognitive workload in MATB-A (M=74.33, SD=14.13) and both MATB-C (M=61.21, SD=16.09); t(9)=3.99, p=0.0016 and MATB-D (M=63.23, SD=13.15); t(9)=2.15, p=0.0302, supporting a significantly higher perceived cognitive workload elicited by the most difficult condition compared to the two easiest conditions. There was also a significant difference between the mean perceived cognitive workload in MATB-B (M=72.15, SD=10.67) and MATB-D (M=63.23, SD=13.15); t(9)=2.01, p=0.0379, supporting a significantly higher workload elicited by the
condition with the second highest difficulty compared to that with the lowest difficulty. No other significant differences were found between means.

Figure 4b.2 The mean perceived level of cognitive workload across participants from the condition with the highest frequency of events (MATB-A) was significantly higher than the conditions representing the lowest frequency of events (MATB-D) and the second lowest frequency of events (MATB-C) at $\alpha = 0.05$. Furthermore, the condition with the second highest frequency of events (MATB-B) had a significantly higher mean than MATB-D ($p<0.05$)

Physiological Indicators

Data from continuous ECG collection were analyzed by individual HRV components and overall HR and subjected to repeated measures ANOVAs. Analysis of the mean HR across conditions, normalized to each individual’s baseline, reveals no significant differences and no general trends.

Comparisons of means via paired t-tests revealed that there was a significant difference between RMSSD values in MATB-A ($M=0.02$, $SD=0.18$) and MATB-C ($M=0.14$, $SD=0.24$); $t(8)=2.79$, $p=0.0117$. This difference supports a significantly lower value for this physiological indicator of stress, RMSSD, corresponding to a higher state of stress, in the most difficult version of the task compared to the second easiest version. Overall, values across all four conditions
follow the expected linear relationship of increasing HRV values with decreasing levels of task difficulty. The lowest HRV values are seen in the most difficult version of the task and the highest HRV values are seen in the easiest version of the task. Comparison of means among pNN50 results across conditions reveals no significant differences (Figure 4b.3).

Figure 4b.3 The difference between normalized mean RMSSD values across conditions was significant (p<0.05), such that MATB-A produced significantly lower RMSSD values on average compared to MATB-D, but pNN50 showed no significant differences across any condition. In both cases, the observed trend aligns with our expectations. Both measures showed a linear increase in HRV component values from MATB-A with the highest frequency of events to MATB-D with the lowest frequency of events.

Paired t-tests to evaluate the differences between SDNN values indicate significant differences between MATB-A (M=-0.11, SD=0.12) and all other conditions [(MATB-B: M=0.02, SD=0.20; t(8)=2.28, p=0.0261); (MATB-C: M=0.00, SD=0.20; t(8)=2.18, p=0.0303); (MATB-D: M=0.09, SD=0.23; t(8)=2.35, p=0.0233)]. These results support that the value for this physiological indicator of stress, SDNN, is significantly lower, corresponding to a higher
state of stress, in the most difficult version of the task compared to all other versions (Figure 4b.4).

![SDNN Values by Condition](image)

**Figure 4b.4** The normalized mean SDNN value in the condition with the highest frequency of events (MATB-A) was significantly lower than that value in the condition with the lowest frequency of events (MATB-D) at $\alpha = 0.05$

**Performance Indicators**

An overall performance score for each version of MFMG was derived (see Analysis for details). Overall performance was analyzed by comparing the means across conditions through repeated measures ANOVA. These analyses revealed no significant differences (Figure 4b.5).
Figure 4b.5 Overall performance scores, grouped by condition. There are no statistically significant differences between the means across conditions, but the expected trend is still largely confirmed. The conditions with the highest frequency of changing events (MATB-A and MATB-B) also have the lowest overall performance score across participants, while the conditions with the lowest frequency of changing events (MATB-C and MATB-D) both have the highest overall performance score across participants.

Discussion

This pilot study demonstrated a convergence of data from various sources, including subjective reports, performance measures, and physiological recordings. Results suggest that there are two disparate versions of the same task which can induce differential levels of physiological and psychological stress.

Although the two subjective stress measures collected did not yield consistent results, I can say with some confidence that the responses via the NASA-TLX are more representative of the cognitive workload I intended to measure, rather than those from the STAI. While the STAI has historically been referred to as a sensitive and reliable measure of anxiety, it is not without its barriers (Marteau & Bekker, 1992). The confidence in our results extends to the frequent and reliable use of the NASA-TLX among medical populations (Alaraj, Tobin, & Birk, 2013; Alkahtani, Aziz, Ahmad, & Darmoul, 2015; L. P. H. Andersen, Klein, Gögenur, & Rosenberg,
2012; Effken, Loeb, Kang, & Lin, 2008; Mazur et al., 2014; Park et al., 2017; Wadhera et al., 2010; Yurko et al., 2010; Zheng et al., 2012), as well as the development and validation of the SURG-TLX (Berg et al., 2015; Hallbeck, Lowndes, & Bingener, 2013; Lowndes, Bingener-Casey, & Hallbeck, 2014; Roy et al., 2015; Weigl et al., 2016; M. R. Wilson et al., 2011), an adaptation specific to surgeons. My future work in this area will address these considerations by administering the short-form version of STAI developed by Marteau and Bekker in 1992, as well using the SURG-TLX scale.

Previous work has called into question the validity of using HR as a metric for performance or stress in surgical populations (Goldman, McDonough, & Roemond, 1972; Payne & Rick, 1986). Additional work has specifically suggested an enhanced sensitivity among HRV measures compared to HR measures to detect acute stress and mental strain among surgeons (Böhm et al., 2001).

Results from the comparison of means for pNN50 did not reach significance at $\alpha = 0.05$, but do follow a linear trend. On the other hand, the trends for SDNN and RMSSD are in the expected direction and comparisons of means do reveal significant differences between conditions. The primary limitation affecting lack of significance in physiological data is likely the small sample size recruited ($n = 10$), but results are still promising.

Data representing overall performance showed a trend similar to that of cognitive workload measured by the NASA-TLX, in which MATB-A and MATB-B versions of the task showed poorer performance and were closer in value to one another, while MATB-C and MATB-D versions of the task showed better performance and were closer in value to one another. The comparisons of means between performance scores across conditions revealed that the differences were not significant at $\alpha = 0.05$, however. This lack of significant findings may
point to the multiple different ways to capture optimal performance on each subtask, and the nuances among the interactions between subtasks.

While performance on the system monitoring subtask was represented by the proportion of correct responses to observed deflections, an alternative metric could have been reaction time. I intentionally chose not to use reaction time due to the nature of the physical set-up of the experiment. With the medical assessment questions displayed on a separate tablet next to the desktop computer, and the remaining three subtasks displayed in one interface and on one computer screen, there was a tendency for participants to focus more on the MATB-II interface when the MATB-II had a higher frequency of changing events, and more on the medical assessment questions when the MATB-II had a lower frequency of changing events. This would result in MATB-A having either very quick reaction times while attending to frequent deflections in system monitoring and many missed medical questions, or many missed deflections in the system monitoring task while participants are attending to the medical questions. On the other hand, MATB-D with a very low frequency of events might encourage more attention to the medical questions on the tablet and a slower reaction time in terms of deflections in the system monitoring subtask. When reaction time was included as the representative performance metric in the systems monitoring subtask, replacing the proportion of correct responses, the difference in overall performance scores across conditions was negligible (data not shown).

Another concern that could have affected performance on the overall MFMG task was the high potential that scores on the medical assessment subtask reflected the proportion of correctly guessed answers rather than the proportion of known answers. As medical students in the midst of training, a majority of participants admitted to guessing answers. To account for this
potential confound, and the potential that some sets of questions may have been easier than others, an overall performance score was calculated excluding the contribution from medical assessment. As a result, overall performance in each condition decreased by roughly the same amount (data not shown), indicating that it was in fact the disparate versions of the MATB-II that contributed almost exclusively to the disparate difficulty of the MFMG task.

The potential for knowledge advancement in the realm of investigations into acute stress interventions in healthcare are enormous as a result of this work. By generating and identifying appropriately stressful versions of what is essentially a “medical MATB-II,” the opportunity to evaluate acute stress using well-established and domain-relevant tasks in a tightly controlled setting can be realized. Investigations into basic processes associated with acute stress and acute stress recovery can be systematically accomplished using this task, which will set the stage for larger-scale investigations into similar processes in an applied setting in the future.

The activation of existing cognitive processes crucial for successful surgical performance, such as multi-tasking, psychomotor control, dynamic resource management, decision-making, uncontrollability, and attention allocation can be accomplished successfully through the use of MATB-II. But to the best of our knowledge, MATB-II has never been used in empirical research to evaluate the stress response in a healthcare setting or among a population of healthcare practitioners. By complementing this task with the added component of memory recall and making it medically-relevant with the addition of medical assessment questions, the total task (MFMG) can more appropriately address additional facets of surgeons’ cognitive processes than either task alone.

Thus, the value in this work has direct real-world applications. We now have an appropriate and domain-relevant, stress-inducing, lab-based task to evaluate the effectiveness of
a stress intervention paradigm in healthcare, under tightly controlled settings. The applicability of a task as valid and reliable as MATB-II into the healthcare setting will extend the possibilities and research avenues within a population of healthcare providers.
Chapter 5
Adherence to auditory coping instructions during an acutely stressful task affects stress and task performance for healthcare providers

Abstract

Healthcare providers often perform under significant stress, during which their performance must be optimal. There is opportunity to take advantage of stress management interventions in this context by capitalizing on sensor technology to acquire physiological data in real time. The objective of this study was to test the influence of visual biofeedback, auditory coping instructions, and emotional intelligence training, and combinations of these interventions on stress and performance while clinicians engaged in a stress-inducing, computer-based task. Participants (N=45) were assigned to one of five groups (N=9 per group), where they completed one high- and one low-stress task. We hypothesized that participants receiving visual biofeedback and auditory coping instructions (Group 3) would present with the lowest self-reported and physiological measures of stress and the highest overall performance. We found that the high-stress task was significantly more subjectively stressful and cognitively demanding, more physiologically stressful according to two electrocardiographic indices, and performance was significantly worse. Results show Group 3 was the only high-stress group reporting significantly higher levels of cognitive demand than its corresponding low-stress group, counter to our expectations. However, it was also one of two groups whose performance on the high-stress task was not significantly worse than performance on the low-stress task. Ultra-short term heart rate variability analysis revealed significantly decreased stress following auditory coping instructions. Taken together, these data show that despite the discrepancy between reported
cognitive workload and performance in Group 3 participants, physiological signs of stress were significantly reduced immediately following adherence to the coping instruction intervention.

Introduction

Heart rate variability-derived biofeedback has a long history (Lehrer, 2013a) as well as a promising future (Lehrer & Vaschillo, 2008), with broad utility. Recent studies show that biofeedback can be effectively used for stress management (Kennedy & Parker, 2018a). A recent systematic review of biofeedback indicated enhanced psychological, physiological, and performance benefits when evaluated among professionals operating in high-stress settings, including soldiers, correctional officers, police officers, professional athletes, senior managers, and post-partum mothers (Kennedy & Parker, 2018a). This finding is encouraging, because restoring stress to optimal levels among these populations can have broad implications for occupational safety and performance. One crucial demographic missing from the biofeedback literature is healthcare practitioners, for whom the consequences of poor performance due to unregulated stress are dire.

Previous work has also suggested that explicit coping instructions advising behavioral adaptations in response to elevated stress levels could provide a meaningful supplement to stress management approaches (Kennedy & Parker, 2018b). Specifically, heart rate variability (HRV) can be used as a proxy for stress and/or cognitive workload (Kennedy & Parker, 2017, 2018b; Wu, Cao, Nguyen, Surmacz, & Hargrove, 2015), particularly in the field of surgery (Dias, Ngo-Howard, Boskovski, Zenati, & Yule, 2018). Given the rapid advancement of modern sensor technology, HRV can now be accurately and reliably captured on ultra-short time scales (Baek, Cho, Cho, & Woo, 2015; Munoz et al., 2015; Nussinovitch et al., 2011; Thong et al., 2003).
Analysis of these ultra-short-term HRV components is sensitive enough to distinguish between mental stress (D. Kim et al., 2008; Salahuddin et al., 2007) and arousal (Schaaff & Adam, 2013) states.

To manage stress effectively in the high-risk domain of healthcare, reliable, unobtrusive measurement and aiding is necessary. The aim of the current study was to evaluate the influence of combinations of biofeedback intervention components. Specific components included: visual biofeedback displayed as heart rate (HR) in a trending format; auditory coping instructions delivered at times of elevated HR; and a brief training video discussing emotional intelligence in relation to coping with stress in a clinical environment.

My research questions and hypotheses were:

1. Do perceptions of stress and cognitive workload differ as a result of different intervention components?
   - Subjective reports of stress and cognitive workload will be lower when participants receive the highest degree of real-time cognitive aid (biofeedback, coping instructions).

2. Do physiological measures of stress differ as a result of different intervention components?
   - Objective indicators of stress will be lower when participants receive the highest degree of real-time cognitive aid.
   - Ultra-short-term objective indicators of stress will be lower immediately following adherence to discrete intervention events (coping instructions).

3. Does performance differ as a result of different intervention components?
- Task performance will be higher when participants receive the highest degree of real-time cognitive aid.

Materials and methods

To capture subtle differences among physiological variables, the current study utilized a validated task, Medically-Focused Multitasking Game (MFMG), in a lab setting to achieve high experimental control (Kennedy & Parker, 2017).

Participants

A total of 45 late-stage medical students (N=32) and residents (N=13) were recruited to participate in this study. Residents were recruited from emergency medicine, surgery, neurosurgery, and obstetrics and gynecology specialties, and from all post-graduate years (1-6). Of the medical students recruited, 23 were enrolled at the end of their final year, and 9 were enrolled at the beginning of their final year. The experiment was approved by and conducted in accordance with the guidelines of the Carilion Clinic Institutional Review Board, and subjects provided written consent before the study began.

Equipment

Electrocardiograph (ECG) signals were collected using a MindWare Technologies LTD (Gahanna, OH) 3-lead ambulatory recording device at a sampling rate of 500 Hz. The ECG signals were acquired and transmitted in real time directly to a PC, viewed only by the researcher, by MindWare’s BioLab Acquisition Software. Three lead-shielded electrode leads attached to disposable ECG electrodes transmitted voltage signals and recorded changes from the
RA (-), LL (+), and RL (ground) locations on the chest. Participant instructions and medical questions were developed using E-Prime stimulation software by Psychology Software Tools, Inc. (Sharpsburg, PA). The auditory coping instruction administered in Group 3 was pre-recorded using a generic voice recorder. The instruction read “Take a deep breath. Count to three out loud as you exhale”, and was administered during the experimental phases when HR exceeded one standard deviation above an individual’s average HR recorded during the baseline period.

MFMG was designed to induce different degrees of stress while recruiting attentional processes and key cognitive processes such as memory recall, and decision-making (Kennedy & Parker, 2017). The MATB-II component of the task (Comstock & Arnegard, 1992) was completed on a 12” Lenovo Yoga 12 laptop computer, while the task instructions and medical questions component of MFMG were displayed on a 21.5” Dell monitor positioned directly above and behind the Yoga laptop. When HR information was available to participants, in Groups 2, 3, and 5, HRV signals were transformed in real time by BioLab and displayed as HR in a trending format and beats per minute, on a 24” Dell monitor placed next to the task display (Figure 5.1).
Figure 5.1 Participant view of the experimental set-up. MATB-II was completed using the joystick, mouse, and the computer screen in the foreground, while medical questions were displayed on the large screen directly above and behind. The screen on the participant’s right displayed HR in Groups 2, 3, and 5

Procedure

Upon arrival, participants reviewed and signed the informed consent. Participants then had time to ask additional questions as the researcher entered their information into the system and compensated them with $20. The researcher then prepared the three sites for electrode attachment by cleaning the skin with alcohol swabs and lightly abrading with NuPrep gel. Once participants were equipped and data was streaming into the BioLab software, participants viewed a 12-minute MATB-II tutorial produced by NASA, overviewing the goals of each subtask and the appropriate controls. The subtasks employed in this study included Resource Management, System Monitoring, and Tracking. Participants in Groups 4 and 5 then watched a 12-minute video describing the tenets of emotional intelligence and their relationship with stress and coping in clinical practice. This video was co-written by all authors and featured one author (PW), a clinician and leading expert in the field.

Following the tutorial(s), participants were allowed to ask additional questions and then began a 20-minute practice phase to get acquainted with MFMG and produce stable levels of performance on MATB-II (Venables & Fairclough, 2009). At the start of this phase of the study, the biofeedback display was initiated for those participants assigned to groups receiving visual biofeedback. Within the first five minutes of the practice phase, participants in Group 3 received one auditory coping instruction. Unlike the medical questions prompted during the experimental phase, questions during this practice phase represented random trivia knowledge, to give participants the opportunity to fully learn the MATB-II subtasks. Participants then reported basic demographic details and responded to form Y-2 of the State Trait Anxiety Inventory for Adults.
(STAI) (Spielberger & Gorsuch, 1983) to indicate their general stress level, a relatively stable trait.

While participants responded to these questionnaires, the researcher calculated average HR and standard deviation, taken from the middle 10-minute segment of the practice phase. The HR value representing one standard deviation above a participant’s average value was used to determine when auditory coping instructions would be administered during ensuing experimental conditions.

Participants were then randomly assigned to one of five experimental groups (Table 5.1). In all groups, the experimental phase involved completing two 10-minute versions of MFMG. One version was considered to be low stress and low difficulty, while the other was considered to be high stress and high difficulty. The order of conditions was counterbalanced across participants within each experimental group.

<table>
<thead>
<tr>
<th>No biofeedback</th>
<th>No prior stress management training</th>
<th>Prior stress management training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biofeedback</td>
<td>Group 1 (N=9)</td>
<td>Group 4 (N=9)</td>
</tr>
<tr>
<td>Augmented biofeedback</td>
<td>Group 3 (N=9)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1 Experimental groups (Biofeedback = continuous visual HR; Augmented biofeedback = continuous visual HR and intermittent auditory coping instructions; Prior stress management training = pre-recorded video covering emotional concepts of self-awareness and self-regulation and their relationship with stress and coping)

Details on the development and specific composition of these tasks were previously published (Kennedy & Parker, 2017). Following each condition, participants reported their self-perceived levels of stress via the short form of Y-1 of the STAI (Marteau & Bekker, 1992), and self-perceived levels of cognitive workload via the NASA Task Load Index (NASA-TLX) scale (Hart & Staveland, 1988). At the conclusion of the study, participants weighed each of the six workload dimensions represented in the NASA-TLX.
Analysis

Analysis of self-report, physiological, and performance data relied on calculating 2 (stress level) x 5 (group assignment) ANOVAs for each data type. Post-hoc Tukey’s HSD calculations were performed to follow up when necessary. Paired Student’s and independent t-tests were also calculated to compare overall differences between high stress and low stress conditions. All results are reported as significant when analyses yielded a p-value < 0.05.

Analysis of physiological signals to derive HRV values was conducted using MindWare Technologies LTD (Gahanna, OH) Heart Rate Variability Analysis Software. All ECGs were visually inspected in the HRV Software and ectopic beats were manually removed. Statistical analysis of all data was then conducted using Microsoft Excel and JMP Pro 14 by SAS Institute, Inc. (Cary, NC).

The segment of data spanning from five to 15 minutes of the 20-minute practice phase was used to represent baseline values for physiological signals and MATB-II performance metrics. This was considered an appropriate baseline period because it intentionally excluded the first five minutes when participants were initially exploring the task, and the final five minutes when participants began to experience fatigue or boredom.

For physiological data, the baseline values were derived on an individual basis, and became the denominator when normalizing data from the experimental phases for each participant (Equation 1).

\[
\frac{\text{Value}_{\text{Experimental}}}{\text{Value}_{\text{Baseline}}}
\]
Values representing MATB-II performance data were calculated separately for each subtask before combining them. For each subtask, a group average and group standard deviation were calculated, after which each participant’s performance on that subtask was transformed into a z-score (Equation 2).

\[
\frac{Value_{individual} - Average_{Total}}{Standard Deviation_{Total}}
\]

To calculate an “overall MATB-II” score, each subtask z-score was summed and divided by the number of subtasks, to assign equal weight given to each subtask. Performance on medical questions was calculated strictly as a percentage and is reported separately from MATB-II performance, because baseline questions were not medically-oriented and thus are not representative of performance on experimental phases.

To determine the impact of the coping instructions on physiological response, additional analysis of participants in Group 3 also involved calculating values for the 10, 20, and 30 seconds preceding and the 10, 20, and 30 seconds following the delivery of an auditory coping instruction. A value representing the change in physiology for each time-domain HRV component was calculated by subtracting the value preceding the stimulus from the value following the stimulus.

Results

Self-report

Independent-samples t-tests were conducted to compare perceived stress in high and low stress conditions. Results show that there was a significant difference in the stress levels reported
in high (M=13, SD=3.82) and low (M=11.07, SD=3.05) stress conditions; t(82)=2.62, p=0.0052 (Figure 5.2).

![Overall State Stress](image)

**Figure 5.2** Self-reported stress state, according to responses on the STAI, was significantly higher across high-stress conditions compared to low-stress conditions (p=0.0052). HS: high-stress, LS: low-stress

Results also reveal a significant difference in perceived cognitive workload between high (M=67.29, SD=17.27) and low (M=49.87, SD= 14.69) stress conditions; t(83)=5.10, p<0.0001 (Figure 5.3).

![Overall Cognitive Workload](image)
Figure 5.3 Self-reported cognitive workload, according to responses on the NASA-TLX, was significantly higher across high-stress conditions compared to low-stress conditions (p<0.0001). HS: high-stress, LS: low-stress.

The 2 (stress level) x 5 (group assignment) ANOVAs revealed no significant differences in perceived stress, according to responses on the STAI. There was, however, a significant effect on perceived cognitive workload; F(9,78)=3.89, p=0.0004. Post hoc comparisons using the Tukey HSD test indicated that the mean cognitive workload reported in the Group 3 high-stress (M=76.18, SD=12.12) condition was significantly higher than that of the Group 1 low-stress (M=46.39, SD=19.57), Group 3 low-stress (M=46.46, SD=17.15), and Group 4 low-stress (M=46.79, SD=5.95). Cognitive workload reported in the Group 5 high-stress (M=71.67, SD=18.10) condition was also significantly higher than in the Group 3 low-stress (M=46.46, SD=17.15) and Group 4 low-stress (M=46.79, SD=5.95) conditions (Figure 5.4).

Figure 5.4 A 2x5 ANOVA revealed interaction effects for self-reported cognitive workload, according to responses on the NASA-TLX, such that Group 3 high-stress had a higher reported cognitive workload than Group 1 low-stress, Group 3 low-stress, and Group 4 low-stress. Group 5 high-stress also had a higher reported cognitive workload than Group 4 low-stress. HS: high-stress, LS: low-stress.
Physiology

Physiological data analysis calculated independent-samples t-tests and 2 (stress level) x 5 (group assignment) ANOVAs for three common time-domain features of HRV and four common frequency-domain features of HRV (Electrophysiology, 1996). The time-domain features analyzed include the root mean square of the successive differences (RMSSD), the standard deviation from normal-to-normal (SDNN), and the percentage of consecutive N-N peaks differing by at least 50 ms (pNN50). Frequency-domain features analyzed include the very low frequency power band (VLF Power; 0.0033-0.04 Hz), low frequency power band (LF Power; 0.04-0.15 Hz), the high frequency power band (HF Power; 0.15-0.40 Hz), and the low frequency to high frequency ratio (LF/HF Ratio).

Results of the independent-samples t-tests comparing values representing entire 10-minute conditions show that there was a significant difference in the SDNN component recorded in high (M=1.02, SD=0.22) and low (M=1.11, SD=0.27) stress conditions; t(78)=-1.71, p=0.0459 (Figure 5.5).
Figure 5.5 Independent-samples t-tests reveal a significantly lower SDNN value on average, indicative of higher stress overall, across high-stress conditions compared to low-stress conditions (p<0.05). Other time-domain components analysed (RMSSD and pNN50) do not indicate a significant different between stress levels. HS: high-stress, LS: low-stress; Norm_RMSSD: normalized root mean square of the successive differences; Norm_SDNN: normalized standard deviation from normal-to-normal; Norm_pNN50: normalized percentage of consecutive N-N peaks differing by at least 50 ms (normalized values calculated according to Equation 1)

Results also reveal a significant difference in the LF Power recorded in high (M=0.96, SD=0.49) and low (M=1.20, SD=0.76) stress conditions; t(70)=-1.68, p=0.0482 (Figure 5.6).

![Frequency-Domain Measures of HRV Across Stress Levels](image)

Figure 5.6 Independent-samples t-tests reveal a significantly lower LF power value on average, indicative of higher stress overall, across high-stress conditions compared to low-stress conditions (p<0.05). Other frequency-domain components analysed (VLF, HF, and LF/HF ratio) do not indicate a significant different between stress levels. HS: high-stress, LS: low-stress; Norm_VLF: normalized very low frequency power band; Norm_LF: normalized low frequency power band; Norm_HF: normalized high frequency power band; Norm_LF/HF Ratio: normalized low frequency to high frequency ratio (normalized values calculated according to Equation 1)

The ANOVAs performed revealed no significant differences for any of the HRV components.
Ultra-short term physiology

Additional analyses of physiological components compared 10-, 20-, and 30-second time windows directly preceding and following coping instruction delivery. This analysis was performed for time-domain components only (RMSSD, SDNN, and pNN50), due to software capabilities. Paired Student’s t-tests were calculated to compare the mean HRV values (RMSSD, SDNN, and pNN50) for each duration (10s, 20s, and 30s) before and after coping instructions. Using the values representing a change in physiology before and after coping instructions for each HRV value, I also conducted 2 (stress level) x 3 (time window) ANOVAs. Significant differences between different time windows representing the same stress level would indicate a lack of sensitivity of the ultra-short time calculation.

**RMSSD**

While there was a significant difference in the RMSSD values observed before (M=43.45, SD=23.54) and after (M=57.80, SD=27.54) coping instruction delivery for the 10-second time windows (t(74)=2.47, p=0.0078), there were no significant differences observed when comparing RMSSD values spanning 20- and 30-second time windows. The 2 (stress level) x 3 (time window) ANOVA indicated a significant effect on RMSSD value; F(5,111)=3.07, p=0.0123. Post hoc comparisons using the Tukey HSD test indicated that the mean change in RMSSD observed in the 10-second time window of the low-stress condition (M=14.42, SD=7.66) differed significantly from that of the 20-second high-stress condition (M=4.82, SD=14.74), the 30-second high-stress time window (M=1.71, SD=10.90), and the 30-second low-stress time window (M=4.71, SD=10.41). Results also show that the 10-second time window of the high-stress condition (M=14.30, SD=21.10) differed significantly from that of the
20-second high-stress condition (M=4.82, SD=14.74), the 30-second high-stress time window (M=1.71, SD=10.90), and the 30-second low-stress time window (M=4.71, SD=10.41).

*SDNN*

Analysis of SDNN values revealed a significant difference before (M=52.55, SD=24.20) and after (M=79.87, SD=32.56) coping instruction delivery for the 10-second time windows (t(70)=-4.21, p<0.0001); before (M=55.13, SD=21.87) and after (M=73.35, SD=22.69) coping instruction delivery for the 20-second time windows (t(76)=-3.61, p=0.0003); and before (M=53.82, SD=19.28) and after (M=69.05, SD=21.55) coping instruction delivery for the 30-second time windows (t(75)=-3.29, p=0.0008); (Figure 5.7).

![SDNN Values from Time Windows Surrounding Coping Instruction Delivery](image)

**Figure 5.7** Ultra-short term HRV analysis reveals a significant increase in SDNN values, indicative of a decrease in stress state, immediately following an auditory coping instruction in Group 3 compared to those values immediately preceding the instruction for 10-second (p<0.0001), 20-second (p<0.05), and 30-second time windows (p<0.05). SDNN: standard deviation from normal-to-normal

No significant effect was detected as a result of calculating the 2 (stress level) x 3 (time window) ANOVA for SDNN values.
**pNN50**

No significant differences between pNN50 values before compared to after coping instruction delivery were detected for any of the time window durations analyzed.

The 2 (stress level) x 3 (time window) ANOVA for pNN50 values did reveal a significant effect on pNN50 values; $F(5,111)=2.72$, $p=0.0236$. Post hoc comparisons using the Tukey HSD test revealed that the mean change in pNN50 observed in the 10-second time window of the low-stress condition ($M=5.28$, $SD=14.02$) differed significantly from that of the 20-second high-stress condition ($M=-4.58$, $SD=16.86$) and the 30-second high-stress condition ($M=-4.97$, $SD=12.76$). Further, the mean change in pNN50 observed in the 10-second time window of the high-stress condition ($M=7.60$, $SD=15.96$) differed significantly from that of the 20-second high-stress condition ($M=-4.58$, $SD=16.86$), the 30-second high-stress condition ($M=-4.97$, $SD=12.76$), the 20-second low-stress condition ($M=-2.96$, $SD=14.39$), and the 30-second low-stress condition ($M=-3.48$, $SD=15.62$).

**Performance**

Performance analysis calculated independent-samples t-tests and 2 (stress level) x 5 (group assignment) ANOVAs for medical question accuracy and the combined MATB-II score. There were no statistical differences observed within the medical questions data. However, an independent-samples t-test revealed significantly higher performance in the low-stress conditions ($M=0.63$, $SD=0.57$) compared to the high-stress conditions ($M=0.11$, $SD=0.67$); $t(82)=-3.91$, $p<0.0001$ (Figure 5.8).
Overall performance on combined MATB-II tasks was significantly lower across the high-stress conditions compared to the low-stress conditions (p<0.0001). HS: high-stress; LS: low-stress

Independent-samples t-tests within individual groups also indicated statistically significant differences across stress levels. In Groups 1, 2, and 5, higher performance was observed in low-stress conditions (Group 1: M=0.63, SD=0.52; Group 2: M=0.70, SD=0.36; Group 5: M=0.81, SD=0.43) compared to high-stress conditions (Group 1: M=0.22, SD=0.43; Group 2: M=0.20, SD=0.35; Group 5: M=0.25, SD=0.42); (Group 1: t(15)=1.78, p=0.0479; Group 2: t(16)=3.00, p=0.0043; Group 5: t(15)=2.65, p=0.0091.

A 2 (stress) x 5 (group assignment ANOVA also revealed a significant effect on performance; F(9,77)=2.65, p=0.0100. The post-hoc Tukey’s HSD test indicated significant differences between the Group 3 high-stress condition (M=-0.31, SD=1.07) compared to Group 2 low-stress (M=0.70, SD=0.36), Group 4 low-stress (M=0.72, SD=0.54), and Group 5 low-stress (M=0.81, SD=0.43); Figure 5.9.
Figure 5.9 Overall MATB-II performance was significantly lower in the Group 3 high-stress condition compared to Group 2 low-stress, Group 4 low-stress, and Group 5 low-stress. HS: high-stress; LS: low-stress

Discussion

Overall these results, though somewhat mixed, support that despite the potential for cognitive overload a multi-modal biofeedback intervention (i.e. Group 3) may introduce according to self-report measures, this specific intervention is still able to positively affect performance and ultra-short term physiological indicators of stress. While the analysis revealed a number of significant relationships, those deserving the most attention include the differences between stress levels in Group 3 specifically. When the results from each outcome measure are considered in tandem for Group 3, an interesting story emerges.
Self-report

Responses to both self-report instruments indicated that the high-stress condition was more stressful and cognitively demanding than the low-stress condition when collapsing all groups. Group 3 participants perceived the high-stress condition to be significantly more cognitively demanding than the low-stress condition, and this was the only group in which this pattern was observed. Notably, if the cognitive aid had been effective in Group 3, we would have expected the higher cognitive demand induced by the high-stress condition to be minimized to the point at which it was indistinguishable from the low-stress condition. Instead, Group 3 was the only condition in which the perception of cognitive demand was significantly higher in the high-stress version of the task. This group received not only visual feedback, similarly to Groups 2 and 5, but also auditory input. The incorporation of the auditory modality was only present in Group 3, and given the perceptions of cognitive workload, this component specifically could have contributed to overload. Previous work has demonstrated the detrimental effect of auditory distractors on workload and performance, including intra-operative phone calls and verbalized patient discomfort (Weigl et al., 2016) as well as music and case-irrelevant conversation (Pluyter et al., 2010). While our goal was to facilitate cognition, it appears that the additional inputs associated with Group 3 may have introduced too many interruptions into task flow.

Physiology

Standard physiological data analysis, using short-term windows of 10 minutes to encompass each complete condition, revealed main effects of stress for two HRV components: SDNN and LF power. In both cases, values were lower in the high-stress condition compared to the low-stress condition.
While the literature broadly supports that the short-term SDNN component of HRV reflects parasympathetic activity (McCraty & Shaffer, 2015; Shaffer, Ginsberg, & Shaffer, 2017), there is much more debate over the appropriate interpretation of LF power (Heathers, 2014). The common misconception is that LF power represents sympathetic activity (Heathers, 2014), yet others have reported primarily vagal influence, particularly during slow breathing or when an individual sighs or takes a deep breath (McCraty & Shaffer, 2015; Shaffer et al., 2017).

The significantly higher SDNN and LF power values observed in the low-stress conditions compared to high-stress conditions in this study corroborate these previous findings on a short-term time scale, supporting the parasympathetic origins of LF power during instructed breathing.

Ultra-short term physiology

Multiple ultra-short term physiological measures were gathered to analyze physiological state immediately before and after coping instruction presentation in Group 3. Since Group 3 was the only condition receiving coping instructions, analysis was limited to participants in this condition. However, only the SDNN measure revealed a consistent trend across time windows, indicating stability and thus lending itself to more reliable interpretation. We can consider these findings to be valid because there is no observed difference when considering these values across different time windows. The magnitude of change in the 10-second time window is not statistically different from that of the 20- or 30-second time windows, reflecting a stability in the SDNN measure across ultra-short time scales.

This finding is compelling and contributes substantially because previous work demonstrates that SDNN is a function of the recording length, such that the longer the recording,
the larger and more stable the SDNN values (Electrophysiology, 1996; Nunan, Sandercock, & Brodie, 2010; Shaffer et al., 2017). Thus we could have expected SDNN values to differ significantly when comparing recording lengths of varying durations. The findings presented here can therefore be considered robust and meaningful. SDNN was the only HRV component analyzed in the ultra-short term time windows that demonstrated significant differences in values before and after coping instruction delivery among Group 3 participants across all durations. At the same time, SDNN values proved to be consistent when comparing values across different durations, unlike RMSSD and pNN50 values which differed significantly based on the time window analyzed.

These data support that the delivery of and adherence to auditory coping instructions has an immediate and beneficial effect on ultra-short term physiological indicators of stress, regardless of the overall stress condition. This immediate physiological benefit is apparent in the data, despite the self-reported high levels of cognitive demand within Group 3. Using SDNN as the index of stress, the findings show that this benefit is maintained over 10-second, 20-second, and 30-second time windows.

Performance

While participants performed significantly worse on the high-stress version of the task compared to the low-stress version of the task across conditions, additional statistical tests reveal more interesting findings. Groups 3 and 4 were the only conditions in which performance was not statistically worse in the high-stress task compared to the low-stress task. This indicates that there was a driving factor in each of these groups contributing to either improved performance
on the high-stress task or impaired performance on the low-stress task, making performance on these tasks indistinguishable.

Performance by Group 3 in the high-stress condition was the only instance of an interaction effect, where performance was significantly worse than other low-stress conditions (in Groups 2, 4, and 5). Performance by Group 3 in the high-stress condition was the only instance of a negative value, and the highest reported variation among participants. Yet performance by Group 3 in the high-stress condition did not differ significantly from that of the corresponding low-stress condition, indicating that the auditory coping instructions and/or visual biofeedback interface contributed to overall performance. The high variability in performance across Group 3 participants could reflect individual differences in responding to the intervention, which could potentially include level of expertise and/or modality. These data seem to suggest that the multi-modal intervention could have been helpful to some participants and unhelpful to others, and in varying degrees. Group 3 consisted of three residents and six fourth-year medical students, raising the question of how much stress the students have been exposed to, and how equipped they might be to recognize and cope with it.

Since performance measures were statistically different between Groups 1, 2, and 5, we can conclude that the intervention for these participants did not contribute to rescuing performance in the high-stress condition. We expected this observation in Group 1, since there were no experimental manipulations introduced. In Group 2, the visual biofeedback seemed to provide no additional benefit compared to the control condition. This suggests that it was the role of the auditory coping instructions that contributed most substantially to performance outcomes in Group 3, despite the high level of cognitive demand it reported induced according to NASA-TLX data. Yet Group 3 also revealed the highest degree of statistical variability among
performance data, in both the high-stress and low-stress conditions. This implies that this intervention design may have been helpful for certain individuals, and harmful for others. Favorable performance outcomes observed in Group 4 could be attributed to the emotional intelligence training, since that is the only factor that distinguished this protocol from that of the control condition. Based on the significantly different performance metrics observed in Group 5 participants, we can again conclude that the visual biofeedback interface provided no additional benefit.

The primary implications associated with these results are the potential benefit of 1) administering coping instructions during an acutely stressful task, despite an increase in perceived cognitive workload, and 2) training individuals on emotional intelligence concepts immediately before beginning an acutely stressful task. The first finding implies that reducing cognitive load, though a worthy and popular endeavor, may not be necessary if the primary goal is to maintain physiology at a baseline level and/or prevent performance decline during acutely stressful scenarios. It is possible that introducing a high level of cognitive support had the unintended effect of distracting or interrupting workflow, but the adaptive behavior initiated by this interruption seemed positive and helpful. It is important to note the variability across performance among Group 3 participants again, which may reconcile these differences. There is the possibility that some individuals found the high level of cognitive aid to be distracting, which ultimately affected their performance, while others may have benefited from the adaptive behavior more substantially. The second finding suggests that although the emotional intelligence tutorial was extremely brief, lasting only approximately 12 minutes, that simply introducing awareness of stress management in a standardized way may be beneficial.
This study had a number of limitations. This evaluation was done with medically familiar participants. Though we targeted recruiting those with surgical experience or interest, some were at an early stage of their career. While we tried to mimic the conditions common to a stressful operation with MFMG, it still likely lacked ecological validity. For the purposes of this work, we felt that a high level of experimental control was necessary. We effectively induced stress according to all measures taken, but the setting and task were still relatively unrelated to the daily workflow of healthcare practitioners.

Future work could parse out individual characteristics to determine the factors driving this variability within conditions. Overall, the opportunities to expand the core concepts of this intervention are promising, especially due to the affordability, timeliness, and unobtrusive approach utilized with sensor technology today. Future work should consider the application of coping instructions, specifically adapted for the domain and workflow, in other high-risk, high-consequence settings. A higher degree of realism and generalizability could be achieved in the healthcare system by more closely approximating clinical care through simulation. Simulation-based training in healthcare can improve performance on clinical tasks while introducing learners to high levels of stress and adverse events, all in a safe environment (Feins et al., 2017). The emphasis on learning to cope appropriately with the stress and adverse events presented ensures that a stress management training program with embedded behavioral coping instructions would be a suitable addition to simulation-based training.

Conclusions

Although the administration of auditory coping instructions may have been perceived as increasing the overall level of cognitive demand associated with an already cognitively
demanding task, following these coping instructions resulted in immediate and overall benefit. Ultra-short term HRV analysis supports that taking a deep breath as a means of responding to increasing stress decreased the physiological indicators of stress triggering the adaptive response. In doing so, overall task performance on an objectively and subjectively more difficult task can be enhanced to approximate the expected performance of a much easier and less stressful version of that task. The high degree of variability in the performance data among Group 3 may indicate that auditory coping instructions may be conducive to some individuals, and detrimental for others. Individual and personality factors that may determine who benefits more were not collected or analyzed in this study, but should be explored in future work.
Chapter 6
Conclusions and Implications

Previous research has demonstrated diminished surgical performance as a result of poorly managed stress, resulting in preventable errors. Managing stress appropriately among this population could thereby enhance surgical performance and patient safety in a meaningful way. Prior work has also shown the beneficial effects of biofeedback on subjective perceptions of stress, objective physiological signs of stress, and task performance across various professions, especially among the many which are considered high-risk and high-consequence (Kennedy & Parker, 2018b). Given the complexities of daily provision of care in operating rooms, surgery is one notable domain that has not been studied, but could likely experience similar levels of benefits from biofeedback interventions for the purpose of stress management. The work described in this dissertation begins to fill this gap by systematically evaluating the influence biofeedback and/or specific coping instructions have on stress and task performance among healthcare professionals completing highly controlled experimental tasks. On a broader scale, the work in this dissertation also contributes to a greater awareness and understanding of the utility of ultra-short term HRV measures to support the immediate benefit of explicit, adaptive coping. My overarching research question was: Does the individual or combined components of biofeedback and coping instructions produce a noticeable benefit for healthcare practitioners? I, broadly speaking, hypothesized that I would observe these benefits as a result of administering continuous biofeedback and/or intermittent coping instructions while participants completed a stress-inducing, computer-based task. The long term goals of this research are to realize decreased perceptions of stress and cognitive workload, decreased physiological signs of cardiovascular stress, and increased overall task performance.
In the final chapter of my dissertation, I will discuss the gaps in research motivating each study, a brief review of the primary results from each study, the limitations encountered with each study, and how these limitations informed the subsequent studies. I will then discuss the implications of this dissertation work as a whole, and how I was able to answer my overarching research question. I will also discuss the specific scientific and practical contribution of my dissertation, and finally I will close by discussing the limitations of this dissertation and future steps in the field that should be considered.

Timing of coping instruction presentation for real-time acute stress management: Long-term implications for improved surgical performance

In order to begin to answer this broad research question and test my hypothesis, I began by first evaluating the potential benefit of coping instructions. Introducing explicit, directive coping instructions in an attempt to alter behavior towards more effective stress management had not been done according to existing literature, so it was important to us to systematically evaluate this component in isolation (i.e. without also introducing biofeedback in the same study). Although the target beneficiary of this work is healthcare providers, I began this process by recruiting a convenience sample of undergraduate students. Similarly, although the target application of such an intervention would be in the healthcare industry while completing surgical tasks, I chose to evaluate these ideas in a lab setting to achieve a higher degree of experimental control. The task used in this study was selected based on a variety of criteria including: a pre-designed task with no alterations or editing required; high visual-spatial and cognitive demands; two distinct pre-made versions eliciting high and low levels of stress; the ability to start and end
the task at any time to ensure equal lengths across conditions; and an overall competitive and stressful nature.

Access to integrated software packages allowed for a flexible and functional experimental design and subsequent analysis. Stimulus presentation software allowed us to display coping instructions at very specific, pre-determined or spontaneous times, and the integration of this software with a stimulus tracker and physiological acquisition and analysis software further allowed us to mark those events. This integration enabled us to gain insight into very specific time-locked events, supporting the analysis of electrocardiographic data on ultra-short time scales. Ultra-short term heart rate variability analysis is something that is relatively new in the field of psychophysiology, but very conducive to my research questions pertaining to this study. I was particularly interested in short-term changes associated with the acute stress response, for which ultra-short term analysis is compatible and accurate.

The results from this study confirmed a significant difference in perceived stress and cognitive workload between the high-stress and low-stress conditions, according to responses on the STAI and NASA-TLX, respectively. Beyond this finding, remaining results were not statistically significant. Compelling trends, however, suggested a physiological benefit when coping instructions are presented at times of elevated HR, and a detriment when they are not presented during times of elevated HR. Additional trends suggest that coping instructions triggered in response to changes in physiology have a more beneficial effect than those administered at random times. A greater overall adherence to following coping instruction was also associated with improved performance in a domain that could be considered quite easy to master, but had little effect on performance associated with extreme difficulty in achieving mastery.
Completing this study provided valuable support to continue investigating the concept of supplemental coping instructions. There were also a number of limitations accompanying this study that went on to inform future phases. Since the task itself was not specifically healthcare-oriented, some of the key cognitive resources that would make such a computer task a tighter corollary to the surgical domain were missing. The primary demands missing were the elements of memory recall, working memory, and multi-tasking. Though the sample recruited was appropriate for the first level of evaluation, we were unable to generalize the findings to healthcare providers, the target users of the intended intervention. Additionally, although the visual coping instructions were intentionally located directly in the line of sight of participants as they were completing the task—on a screen above and behind the computer they were playing on—participants may not have noticed or registered the instruction in many cases. This could have been a byproduct of perceptual narrowing which was likely to occur during such a high-stress, visually demanding task. Finally, though the experimental design was comprehensive and consisted of a within-subjects design, I noticed a marked fatigue effect by the final condition, regardless of the random condition assignment and order of conditions.

As I prepared for future studies, these limitations were at the forefront of the planning process. The following considerations to alter my future experiments were evident: I needed to design a task that could capture specific types of cognitive processing I wanted to elicit; I needed to recruit healthcare practitioners in order to generalize any findings; I needed to reconsider information presentation format and modalities; I needed to reevaluate the experimental protocol to minimize the fatigue effect and to ensure I could recruit the target population; and I needed to establish that I was introducing components of an intervention that would actually be useful to the intended population of users.
Contextual Inquiry and Analysis

While there were a number of critical experimental and logistical changes to consider, the most urgent consideration was in addressing the utility of such an intervention to healthcare providers themselves. Evidence for the effectiveness of a biofeedback intervention among healthcare practitioners did not exist in the literature prior to this work, despite its demonstrated effectiveness in other high-stress professions (Kennedy & Parker, 2018a). Results from the systematic review reported in Chapter 2 also supported the benefit of tailoring interventions to the demographic of interest to achieve the greatest benefit. The first study conducted in this dissertation (Chapter 3) suggest that pursuing these concepts, though unprecedented, would be a worthy and promising endeavor (Kennedy & Parker, 2018b). Without confirming that this demographic would welcome my ideas, and without explicitly acknowledging the importance of including the end users in the design process to facilitate its use in their context, there would not be enough evidence to suggest this work would be meaningful for its intended application. In light of this, before I could move on to future experimental evaluations, I conducted face-to-face contextual inquiry interviews with attending surgeons to gauge their interest and support, and to get a better idea of specific design features and barriers that are important in their context.

The main findings from the subsequent contextual analysis aligned with my expectations in some ways, and provided new information in others. As expected, interviewees corroborated that the modality of information delivery was incredibly important in the context of surgery. One physician specifically mentioned that if information about his physiological state was available to the rest of the surgical team, and his team could therefore observe his stress signature, he would be humiliated. This line of reasoning reinforces the previous finding that admitting one’s
stress would be perceived to be a weakness or a sign of incompetency. In terms of modality, other interviewees emphasized the excessive visual and auditory inputs associated with their current operating room environment, suggesting that additional visual information could be overwhelming or distracting. If visual information was available, however, most surgeons agreed that the format of that information is critical. Many of the interviewees offered that respiratory rate or heart rate could be useful, and especially if they were displayed in a format similar to the patient vitals they are used to observing during surgeries. When stress management was discussed, almost everyone interviewed mentioned focusing on breathing as an intuitive coping mechanism. The final discussion point that was almost unanimous was the utility an intervention such as biofeedback could have on trainees who are still learning on a day to day basis how to deal with their stress. Some physicians mentioned incorporating biofeedback and coping training into simulations for residents and fellows.

The results of the study informed the subsequent design and implementation of the studies that followed. One key consideration was that the modality of biofeedback and coping instructions could differ to avoid overload through one channel. Other results supported the specific type and format of information physicians considered to be potentially helpful in managing their stress. Though I did not specifically mention the idea of introducing explicit coping instructions into a biofeedback intervention during the interviews, the suggestion that focusing on breathing is an inherent mechanism of coping for many of the attendings confirmed that breathing should continue to be the primary component of the coping instruction moving forward. Lastly, feedback regarding who would benefit from the stress management intervention reframed the long-term application of this work for me. The implications shifted more towards training novice surgeons in particular, and away from surgeons in general.
Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task

In order to advance to the final phase of systematic evaluation of biofeedback alone and in combination with coping instructions, the next requirement was to identify a task capable of meeting all of the experimental needs. I sought a task that was relevant for healthcare practitioners, induced dynamic decision making, memory recall, working memory, attention processes, and multitasking. I envisioned a task that could incorporate all of these cognitive processes, while also simultaneously inducing differential levels of stress. Since it was also a priority to recruit residents and fellows whose time is valuable and scarce, and I was well aware of the potential for fatigue effect associated with a longer experimental protocol, I focused on condensing the task and procedure into an appropriate timeline. The purpose of this next phase of my dissertation work was thus to identify a task with all of these features, adapt and iterate this task if necessary, and validate its role in inducing high and low levels of self-reported stress, physiological stress, and task performance. This phase did not involve introducing any experimental manipulations, such as biofeedback or coping instructions, but to derive an appropriate task in preparation for systematic evaluation later.

By consulting the literature surrounding experimental, lab-based tasks that induce appropriate levels of stress to detect a difference from baseline or between stress levels, I started by identifying the Trier Social Stress Test (TSST), the Multi-Attribute Task Battery (MATB-II), and some homegrown options as the main contenders. Despite the broad utility of TSST to induce stress in a controlled, lab setting, I encountered some concerns. The main concerns were in eliciting the right kind of stress, and ensuring that it would be logistically possible to conduct.
TSST is effective in inducing stress, but it is typically characterized as psycho-social stress, which is not quite the same as the stress experienced by surgeons in the operating room. This option would also require extensive development and involvement on the part of subject-matter experts, which was identified as a meaningful barrier. The appeal of using the MATB-II task developed by NASA was its higher level of experimental control, its effectiveness in inducing and accurately measuring cognitive workload in other high-risk, high-consequence practitioners, the broad range of cognitive processes required to successfully perform the entire task, and its key element of uncontrollability. The main hesitation was that MATB-II failed to address working memory or memory recall processes. The ability to tailor the task to my own needs through code provided by NASA, and the ease of supplementing this task with memory components such as multiple-choice questions reconciled the weaknesses of MATB-II in this context. While making decisions regarding the task and updating my model, I continued to receive feedback from healthcare professionals (for insight into appropriate stressors) and medical curriculum developers (for insight into medical questions to include).

Following this extensive development process, I evaluated its effectiveness in inducing appropriate levels of stress and performance differences through a validation study. This validation process supported that the new task, Medically-Focused Multitasking Game, was appropriate for this demographic. Overall, validating MFMG achieved what I hypothesized it would: it produced one version of difficulty that was perceived as significantly more stressful and cognitively demanding, which was also more objectively stressful according to physiological measures. Unlike the self-report questionnaires used in the first study, which were lengthy and time-consuming, the questionnaires used in the MFMG validation study were much more
practical. Along with being practical and saving a good deal of time, these questionnaires still proved to be sensitive enough to measure the constructs of stress and cognitive workload.

The primary limitation discovered as a result of completing this study was the small sample size, which likely contributed in large part to the lack of statistically significant differences among performance measures across the four versions of the task. While this could also potentially be attributed to the shortened length of each version of the task, which were only 5 minutes each, this was still a major concern. Despite the lack of significance detected among performance measures, however, this study did effectively generate two versions of the task that elicited significantly different subjective and objective indicators of stress, the primary need in future studies.

Adherence to auditory coping instructions during an acutely stressful task affects stress and task performance for healthcare providers

With successful validation of a stress-inducing, lab-based task that would likely accomplish my scientific research goals, the next considerations involved designing the experiment itself. In response to the concern about not detecting performance differences over short time scales, I increased the condition length to 10 minutes each. Keeping the potential for fatigue effect in mind, I decided to limit participation to the completion of two 10-minute tasks—one high stress and one low stress. This changed the way I designed the groups for this final study, requiring each participant to be placed into only one group and receive only one type of experimental manipulation. Ultimately, these decisions affected my subject recruitment plan by requiring a larger sample size to arrive at equal sizes in each of the 5 groups. To overcome the challenge of recruiting enough participants at a small medical center, I focused on residents in
the specialties of emergency medicine, surgery, neurosurgery, and obstetrics and gynecology at Carilion Clinic, as well as final-year medical students from Virginia Tech Carilion School of Medicine and Edward Via College of Osteopathic Medicine.

Other logistical considerations surrounded working through equipment challenges to enable the full and flexible functionality of collecting physiological data, presenting visual biofeedback, presenting auditory coping instructions, and setting up the task itself. In determining the timing and format of coping instructions, I used evidence from the first study to inform my approach here. Since physiologically-relevant instructions held the most promise and visual instructions were largely ignored, I chose to deliver auditory coping instructions at moments of heightened physiological stress.

Beyond biofeedback and coping instructions as isolated entities, this study also incorporated the tenets of emotional intelligence theory relevant to stress management—namely, self-awareness and self-recognition. Biofeedback could be considered a means of self-awareness, because internal states that might otherwise escape conscious awareness were being externalized, which theoretically could eliminate the need to be self-aware on one’s own. Furthermore, following coping instructions could be considered a means of self-regulation, because behavioral adaptations were being off-loaded from the user’s cognition and instead being externalized, which theoretically could eliminate the need to self-regulate on one’s own. With these concepts in mind, I worked with an expert in emotional intelligence, Chief Medical Officer of Carilion Clinic, Dr. Patrice Weiss, to develop a brief training video explaining how physicians can apply emotional intelligence concepts and practice to coping under stress. Though brief, this video was powerful due to its specificity in context, audience, and setting. Each participant had the opportunity to hear about how stress affects clinical performance, and learned specific behavioral
coping mechanisms from an expert in the field who also holds a critical leadership role in the professional lives of the participants enrolled. Beyond the personalized influence this video may have had, it also had the benefit of integrating all of these crucial concepts into a standardized form. Every participant viewed the same video and received the same training, regardless of any prior familiarity individuals may have had with some of the emotional intelligence concepts. The purpose of this feature of the experimental protocol was to investigate whether calling attention to one’s self-awareness (with biofeedback) and providing support on how to self-regulate under stress (with coping instructions) could complement my approach. Further, with the incorporation of the video Dr. Weiss recorded, I could ask whether watching a brief overview of these stress management concepts could replace the need for the biofeedback training I am proposing.

Analysis of the data produced from this study revealed that the task produced sufficient levels of stress, resulting in significantly higher self-reported stress and cognitive workload, higher physiological indicators of stress, and worse performance in the high-stress condition compared to the low-stress condition. However, when considering differences within individual groups, the group receiving the most cognitive support—via visual biofeedback and auditory coping instructions—was only one of two groups, along with Group 4, whose performance on the high-stress and low-stress conditions was indistinguishable. All remaining groups performed significantly worse in the high-stress condition compared to the low-stress condition. Finally, ultra-short term physiological data analysis revealed that physiological indicators of stress immediately following coping instruction delivery were significantly reduced following, compared to the time immediately preceding the instruction. These findings suggest that receiving and following explicit coping instructions can positively influence overall task performance and immediate physiological stress.
Although findings from my initial work described in Chapter 3 suggest support for RMSSD as an indicator of stress on short and ultra-short time scales, the findings described in Chapter 5 fail to support the same conclusion. In this case, SDNN proved to be a more valuable measure of stress. Due to analytical constraints in the initial study, RMSSD analysis was more accessible. For this reason I focused almost exclusively on RMSSD rather than exploring all HRV components, and thus I did not evaluate SDNN comprehensively in the study outlined in Chapter 3. With additional resources available, I would expect to have observed similar trends in SDNN revealed through RMSSD analysis.

Generally speaking, the results of the study explored in Chapter 5 provided support for many of my hypotheses, but introduced expected and unexpected limitations. The main findings suggest that although brief emotional intelligence training and biofeedback presentation seem to have little effect on the measures I collected, the presence of auditory coping instructions did influence overall performance and physiological responses. Analysis revealed that by considering only one slice of the data collected, the findings are inconclusive. It was necessary to take not only a comprehensive look at the data available to reconcile the discrepancy between self-report and physiological data, but also a more innovative and sensitive approach to tease out the differences within physiological data where they existed.

Despite deliberate attempts to design the most appropriate experimental protocol and method, I was still met with a number of limitations in the final phase of my dissertation work that should be considered in future work dealing with similar concepts. One concern was relating to the medical questions prompted. Although these questions were approved by a fourth-year medical student and medical school faculty members, they spanned many specialties and levels of expertise, making high performance difficult for most participants. This was the case
regardless of whether the participants were under high levels of stress or not. Additionally, even though I tried to compensate for the lack of ecological validity by specifically recruiting similar cognitive processes involved in a healthcare practitioner’s daily duties, the task did not represent the highest fidelity option, and thus could possibly induce a reduced level of stress for this population compared to a higher fidelity simulation. Finally, while capturing the insights gained through this study was important and informative with implications for individual practitioners, the potential benefit of these data could be even greater if we considered the dynamics among team members as well. Patient care ultimately manifests from the interactive nature of individual providers working on a team, and is affected by individual and shared experiences. In the same way, stress is a factor that influences not only the mindset and performance potential of individual providers, but of the entire team. Incorporating the team as a whole into future study could elucidate interactive, team-based dynamics that might otherwise go undetected when focusing on individuals alone.

With these limitations in mind, I have to next consider how to continue along this line of work in a more advanced way to overcome these barriers. Tailoring the medical questions to each individual’s specialty and/or level of expertise could ensure that this task is more widely applicable and accessible, to enhance generalizability of findings. In order to increase the ecological validity and fidelity of these experiments, the next natural step would be to introduce these concepts into simulated scenarios. Not only would this increase the realism of the dynamics I am interested in replicating, but it would also enable a greater degree of technological integration with realistic interfaces and current practice. As I learned from looking at the bigger picture, we will always miss the potential influence the rest of the surgical team may have on these processes if we do not incorporate the entire team. By evaluating only one member of the
team, we fail to capture inter-personal and social dynamics that have the potential to influence cognitive, physiological, and behavioral changes. Expanding future research to teams would also further enhance the realism of the experimental design and the generalizability of findings.

Implications

The individual components of this research have each contributed to advancing our knowledge of how biofeedback affects various measures, and this has been accomplished by relying on similar approaches throughout this process. Each of the studies described took the form of tightly controlled lab studies and incorporated experimental psychology approaches. With this approach comes a low level of fidelity, mundane validity, but a high degree of experimental control. Thus when differences are detected among the variables of interest, we can appropriately relate these differences to deliberate manipulations introduced into the study design. Each of these studies also have in common the use of non-invasive sensors to detect physiological changes, and the corresponding methodological and analytical emphases. Specifically, physiological data analysis using ultra-short, time-locked calculations surrounding discrete events constituted the crux of the analytical methodology. Considering the focus on experimental and laboratory approaches, this work has enormous translational impact, with the potential to influence clinical practice and care. I have begun to fill a gap by contributing to the foundation of a potential intervention to be explored and refined further, and perhaps implemented in applied settings. Despite the limitations encountered in each of these studies, trending and significant results went on to inform subsequent phases of the dissertation work described (Table 6.1).
General results

**Chapter 4a**
Attending physicians anticipate value in a biofeedback intervention geared towards training residents to cope under stress, but have concerns regarding modality and format of information presentation, as well as the potential to induce additional cognitive workload.

**Chapter 4b**
A computer-based task can be used to induce different levels of stress among a population of medical students, corresponding to significant differences in perceptions and physiological signs of stress, and non-significant differences in overall task performance.

<table>
<thead>
<tr>
<th>Perception</th>
<th>Physiology</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter 2</strong></td>
<td>Biofeedback improved or had no effect on perceived stress</td>
<td>Biofeedback improved or had no effect on perceived stress</td>
</tr>
<tr>
<td><strong>Chapter 3</strong></td>
<td>Visual coping instructions did not affect perceived stress or cognitive workload</td>
<td>Visual coping instructions improved ultra-short term measures of heart rate variability (not significant)</td>
</tr>
<tr>
<td><strong>Chapter 5</strong></td>
<td>Auditory coping instructions increased perceived cognitive workload</td>
<td>Auditory coping instructions improved ultra-short term measures of heart rate variability</td>
</tr>
</tbody>
</table>

Table 6.1 Pilot studies described in Chapters 4a and 4b yielded results supporting further investigation into the role of biofeedback and explicit coping instructions on stress management. Results from the systematic literature review in Chapter 2 suggest that biofeedback never induced negative effects on outcome measures, but rather either induced positive effects or had no effect at all. The experiment described in Chapter 3 suggests the same with visual coping instructions. The final experiment, outlined in Chapter 4, indicates that auditory coping instructions did induce higher cognitive workload, but still improved remaining outcome measures.

Given the results outlined in Table 6.1, we can extrapolate meaning and larger implications that extend beyond the lab. These data suggest that we can acquire robust physiological measures that reflect sensitive, reliable, real-time changes in physiological state, as well as in cognitive state and performance. I have shown support for the neurovisceral integration model by exposing an intricate relationship between cardiovascular changes, neural functioning associated with the prefrontal cortex and other key areas, and behavioral adaptations via active coping. The data at large support the role of coping instructions, whether independently or in conjunction with biofeedback, among a population of healthcare providers. It is clear when comparing results across each component of this dissertation work that modality of information presentation matters; it matters to the subject matter experts, and it also matters in
terms of overall results. In the final study reported, I altered the modality of coping instructions, based on some of the limitations associated with the first study and suggestions from target end users, and this change in modality seems to have had an effect. While results were trending in the first study, we can theorize that they did not reach significance due to the impact of perceptual narrowing, which potentially prevented individuals from registering and responding to the visual coping instructions. I also noticed that the visual biofeedback administered in the final study had very little to no effect, perhaps for the same reasons. But in this study, when coping instructions were delivered through the auditory channel, there was in fact an observed difference. The increased cognitive workload perceived in this study could be indicative of cognitive overload due to excessive inputs, reflective of a potential increase in germane load, according to the Cognitive Load Theory. Despite increasing the level of germane load experienced, I observed a corresponding adjustment within other categories simultaneously. These findings in conjunction could also suggest that the explicit knowledge of one’s physiological state may not be necessary in order to manage one’s stress, but rather that active coping is influential.

Scientific and Practical Contributions

The scientific conversation surrounding applied uses of biofeedback has been expanded as a result of this dissertation work. I have shown through this work that using these fine-grained methods and analyses can fill a practical gap to applied clinical care. By reviewing the literature in reference to using biofeedback for stress management purposes, I have elucidated a nuance that had been previously unrecognized—while biofeedback is effective in many ways for general populations, its effectiveness is more comprehensive and notable in groups in high-risk
professions. This is particularly true when the intervention is tailored to the context and specific performance outcomes of each group.

The experiments I conducted that contribute to this dissertation also provide further evidence that: biofeedback would be embraced by attending physicians as a training mechanism for early stage residents to enhance their stress management ability; a highly controlled environment and appropriate task can induce sufficient levels of stress among healthcare practitioners; a rudimentary biofeedback intervention can have significant implications for the way stress is perceived and physiologically manifested, and can influence performance; and coping instructions are a viable option to address stress management without introducing lengthy training protocols.

This final implication has the potential to produce substantial contributions to clinical training and practice. Introducing an effective stress management training option for residents in the early stage of practice could have enormous implications for future generations’ capacity to perform optimally under stress. In doing so, this approach could have a positive influence on minimizing errors and adverse events, thereby enhancing patient safety. Considering the minimal time required to train individuals—I have shown the effectiveness of introducing coping instructions at physiologically relevant times alone, without the need for biofeedback, biofeedback training, or emotional intelligence training—the enormous potential benefit would be accompanied by negligible potential costs. Residents would not be required to take time away from learning or clinical practice in order to train extensively. Furthermore, the financial cost and amount of effort required to deliver the intervention would be relatively trivial.
Future Steps

Overall, each phase of this work has moved us closer towards realizing the overall goal of improving performance under stress for healthcare professionals. I have effectively begun to fill some of the gaps surrounding using biofeedback as a stress management intervention for physicians, utilizing physiological signals for data acquisition and analysis in this setting, and conveying the effectiveness of the proposed coping instructions included in this work. In order to continue along this trajectory, the next appropriate steps to continue this conversation primarily include: evaluating these interactions during higher fidelity simulations; recruiting multiple individuals comprising a surgical team to participate at once; systematically evaluating alternative modalities of information presentation, such as visuohaptic indicators; and incorporating additional physiological inputs to enhance the ability to accurately estimate measures of cognitive workload in real time.

Specific future research questions rely on the ability to detect cognitive overload with physiological sensors so the user (a healthcare practitioner) can be notified when performance is at risk, and cope accordingly. The future overarching research question is: what is the optimal design for, and development of, physiologically-based aiding? Corresponding hypotheses are: 1) auditory and/or haptic feedback will be the most effective channel to deliver information through; 2) accurate estimation of physiological, cognitive, and performance states will be achieved by incorporating multi-modal measures, such as galvanic skin response, accelerometry, skin temperature, and heart rate variability, in combination; 3) though cognitive load may be considered high in conditions of substantial cognitive support, the mental state will not be indicative of overload, reflected by a physiological and performance maintenance; and 4) when designed appropriately for the demographic and context, biofeedback and/or coping
interventions will improve or maintain outcome measures. Achieving this variety of future work could produce a robust, context-specific, and sensitive system capable of optimizing surgical performance. The work described in this dissertation prepare me to pursue these future questions by establishing the utility of biofeedback and coping instructions for stress management in novice physicians and medical students. This utility was established not only according to subject matter experts’ perceptions, but also a lineage of experimental evidence.
References


Payne, R. L., & Rick, J. T. (1986). Heart Rate As an Indicator of Stress in Surgeons and


Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the


Appendices

Appendix A: Publications and presentations of dissertation work

Publications


Invited talks


Podium presentations


Kennedy, L., Parker, S. H. Making MATB-II medical: Pilot testing results to determine a novel


Poster presentations


Appendix B: Self-report instruments

State Trait Anxiety Inventory for Adults, form Y-1

SELF-EVALUATION QUESTIONNAIRE STAI Form Y-1

Please provide the following information:

Name ___________________________ Date ____________ S ______

Age __________ Gender (Circle) M F T ______

DIRECTIONS:
A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement to indicate how you feel right now, that is, at this moment. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

1. I feel calm ......................................................... 1 2 3 4
2. I feel secure ....................................................... 1 2 3 4
3. I am tense ......................................................... 1 2 3 4
4. I feel strained .................................................... 1 2 3 4
5. I feel at ease ..................................................... 1 2 3 4
6. I feel upset ....................................................... 1 2 3 4
7. I am presently worrying over possible misfortunes ..... 1 2 3 4
8. I feel satisfied ................................................... 1 2 3 4
9. I feel frightened .................................................. 1 2 3 4
10. I feel comfortable ............................................. 1 2 3 4
11. I feel self-confident ........................................... 1 2 3 4
12. I feel nervous .................................................. 1 2 3 4
13. I am jittery ....................................................... 1 2 3 4
14. I feel indecisive ............................................... 1 2 3 4
15. I am relaxed ................................................... 1 2 3 4
16. I feel content .................................................. 1 2 3 4
17. I am worried .................................................. 1 2 3 4
18. I feel confused ................................................ 1 2 3 4
19. I feel steady ................................................... 1 2 3 4
20. I feel pleasant ................................................. 1 2 3 4
State Trait Anxiety Inventory for Adults, form Y-2

SELF-EVALUATION QUESTIONNAIRE
STAI Form Y-2

Name __________________________ Date ____________

DIRECTIONS
A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement to indicate how you generally feel. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe how you generally feel.

21. I feel pleasant ................................................................. 1 2 3 4

22. I feel nervous and restless .................................................. 1 2 3 4

23. I feel satisfied with myself .................................................. 1 2 3 4

24. I wish I could be as happy as others seem to be .................... 1 2 3 4

25. I feel like a failure ............................................................ 1 2 3 4

26. I feel rested ..................................................................... 1 2 3 4

27. I am "calm, cool, and collected" ......................................... 1 2 3 4

28. I feel that difficulties are piling up so that I cannot overcome them ........................................ 1 2 3 4

29. I worry too much over something that really doesn't matter ......................................................... 1 2 3 4

30. I am happy ..................................................................... 1 2 3 4

31. I have disturbing thoughts .................................................. 1 2 3 4

32. I lack self-confidence ......................................................... 1 2 3 4

33. I feel secure .................................................................... 1 2 3 4

34. I make decisions easily ....................................................... 1 2 3 4

35. I feel inadequate .............................................................. 1 2 3 4

36. I am content .................................................................. 1 2 3 4

37. Some unimportant thought runs through my mind and bothers me .............................................. 1 2 3 4

38. I take disappointments so keenly that I can't put them out of my mind ........................................... 1 2 3 4

39. I am a steady person ......................................................... 1 2 3 4

40. I get in a state of tension or turmoil as I think over my recent concerns and interests ..................... 1 2 3 4
State Trait Anxiety Inventory for Adults, Short Form

Appendix A: Self-evaluation questionnaire (Y-6 item)

Name ........................................... Date ..............

A number of statements which people have used to describe themselves are given below. Read each statement and then circle the most appropriate number to the right of the statement to indicate how you feel right now, at this moment. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Somewhat</th>
<th>Moderately</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I feel calm</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2. I am tense</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3. I feel upset</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4. I am relaxed</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5. I feel content</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6. I am worried</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Please make sure that you have answered all the questions.
NASA Task Load Index

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

<table>
<thead>
<tr>
<th>Name</th>
<th>Task</th>
<th>Date</th>
</tr>
</thead>
</table>

**Mental Demand**
How mentally demanding was the task?

- Very Low
- Very High

**Physical Demand**
How physically demanding was the task?

- Very Low
- Very High

**Temporal Demand**
How hurried or rushed was the pace of the task?

- Very Low
- Very High

**Performance**
How successful were you in accomplishing what you were asked to do?

- Perfect
- Failure

**Effort**
How hard did you have to work to accomplish your level of performance?

- Very Low
- Very High

**Frustration**
How insecure, discouraged, irritated, stressed, and annoyed were you?

- Very Low
- Very High
Appendix C: Contextual Inquiry interview questions

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are there ever times during surgeries when you feel like things are</td>
</tr>
<tr>
<td>outside of your control? Times when, even if you had all the resources</td>
</tr>
<tr>
<td>and time in the world, the nature of the situation is overwhelming?</td>
</tr>
<tr>
<td>How often would you say you feel this way?</td>
</tr>
<tr>
<td>Would it be fair to call such a situation “stressful”?</td>
</tr>
<tr>
<td>Can you share a specific story that illustrates this? Regardless of</td>
</tr>
<tr>
<td>whether the outcome was positive or negative.</td>
</tr>
<tr>
<td>How do you deal with this and/or restore control?</td>
</tr>
<tr>
<td>Thinking back on the story you shared, during time-sensitive and serious</td>
</tr>
<tr>
<td>(stressful) procedures, what sorts of things do you pay the most</td>
</tr>
<tr>
<td>attention to?</td>
</tr>
<tr>
<td>Where do you allocate visual attention? What do you listen for the most</td>
</tr>
<tr>
<td>in the environment? What types of haptic/touch sensations do you</td>
</tr>
<tr>
<td>attend to the most?</td>
</tr>
<tr>
<td>Are you aware of any of your own physiological changes that take place</td>
</tr>
<tr>
<td>that you’re aware of at times like this?</td>
</tr>
<tr>
<td>How do you address these physiological changes, if at all?</td>
</tr>
<tr>
<td>Can you think of a way this could be externally represented that</td>
</tr>
<tr>
<td>might make this process easier/more reliable for you?</td>
</tr>
<tr>
<td>Imagine a system that could convey biometric information.</td>
</tr>
<tr>
<td>Can you describe what this might look like to you?</td>
</tr>
<tr>
<td>What information would be helpful to you?</td>
</tr>
<tr>
<td>How would you imagine it should be presented?</td>
</tr>
<tr>
<td>When would this information be most helpful?</td>
</tr>
<tr>
<td>What might some foreseeable barriers to this system be from your</td>
</tr>
<tr>
<td>perspective?</td>
</tr>
</tbody>
</table>
# Appendix D: Heart rate variability measures

<table>
<thead>
<tr>
<th>Variables</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart rate</td>
<td>Average heartbeats per minute over a time window</td>
</tr>
<tr>
<td>R-R intervals</td>
<td>Variability among R-R intervals from ECG data provide heart rate</td>
</tr>
<tr>
<td>Low frequency power</td>
<td>HRV power spectrum ranging from 0.04 to 0.15 Hz</td>
</tr>
<tr>
<td>High frequency power</td>
<td>HRV power spectrum ranging from 0.15 to 0.4 Hz</td>
</tr>
<tr>
<td>Low frequency/High frequency ratio</td>
<td>Ratio of low frequency power to high frequency power</td>
</tr>
<tr>
<td>Total frequency power</td>
<td>HRV power spectrum ranging from 0.005 to 0.4 Hz</td>
</tr>
<tr>
<td>Root mean square of the successive differences</td>
<td>HRV time-domain measure calculating the variation among successive R-R intervals</td>
</tr>
<tr>
<td>Standard deviation from normal-to-normal</td>
<td>HRV time-domain measure calculating the variation of R-R interval values in reference to a mean</td>
</tr>
<tr>
<td>pNN50</td>
<td>HRV time-domain measure calculating the percentage of successive R-R intervals that differ by 50 ms or more</td>
</tr>
<tr>
<td>Coherence index</td>
<td>Primary marker of psychophysiological coherence state, derived from the degree of symmetry of the HRV waveform; a higher value indicates higher symmetry and a more positive state</td>
</tr>
</tbody>
</table>
Appendix E: Correlations among outcome measures in Chapter 5
Calculating correlations among all variables collected in Chapter 5, which includes a comprehensive list of HRV components, MATB-II performance, and subjective indicators of stress (STAI) and cognitive workload (NASA-TLX) illustrate expected significant relationships. Correlations appearing in blue text represent significantly positive relationships while correlations appearing in red text represent significantly negative relationships, at a p-value of 0.05 in all cases. RMSSD, a well-established HRV component reflective of known parasympathetic control, has a significant positive correlation with other known parasympathetic measures, such as SDNN (0.5547), pNN50 (0.7466), HF/RSA Power (0.8412) and a significant negative correlation with known sympathetic measures, such as HR (-0.5882) and LF/HF Ratio (-0.2400). These data also support that NASA-TLX and STAI measures are significantly positively correlated. However, they do not reveal significant correlations between NASA-TLX and STAI scores compared to other outcome measures, or task performance on the MATB-II portions of the task and any other outcome measure.
Appendix F: Alternative representation of Figure 2.1

An alternative theoretical framework for the relationship between stress (shown here as cognitive workload) and performance posits an acceptable range of cognitive workload to ensure the highest level of performance. While this representation still indicates the optimal level of cognitive workload desired to reach highest performance (the vertex of the parabola shown above), it also acknowledges an area under the curve, suggesting a more fluid and less abrupt transition between unfavorable and favorable levels of stress or cognitive workload (Figure from Zenati, 2018).