2020

Comparative Examination of the Empatica E4 to Record Heart Rate Variability Metrics

Andres Rosero
University of Central Florida

COMPARATIVE EXAMINATION OF THE EMPATICA E4 TO RECORD HEART RATE VARIABILITY METRICS

by

ANDRES ROSERO

A thesis submitted in partial fulfillment of the requirements
for the Honors degree of Bachelor of Science
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2020

Abstract

The increased accessibility of cyber technology has resulted in advancements in international communications and information sharing never seen in human history. With this new age of digital software comes the proliferation of illegal online activity and cyber terrorism. Repercussions of cyber-attacks have ranged from identity theft to leaks of classified state secrets. To combat this threat, the Department of Defense (DoD) established the Cyber Mission Force (CMF) to head operations in the interests of protecting against cyber-attacks. One of the CMF's initial projects involves the creation of a Performance Assessment Suite (PAS), a training program designed to improve the training of cyber team members via modeling behaviors and physiological data. One of the primary objectives of the PAS is to evaluate the efficacy of select physiological recording equipment in order to implement in cyber training missions. This project serves to determine the viability of the Empatica E4 as an HRV recording device by comparing its quality of data to another, reliable data collection device – the Equivital EQ02 by leveraging a project involving resonance breathing training with police cadets. The results of this project determined that the E4 was unable to compare favorably for some time domain indices to the EQ02 but did have some slight similarities in data with broader time domain metrics.

Table of Contents

Introduction

Cyberterrorism is a growing threat to many who have an increasing amount of sensitive private information online. Repercussions of cyber-attacks have ranged from identity theft to leaks of classified state secrets. To combat this threat, the Department of Defense (DoD) established the Cyber Mission Force (CMF) to head operations in the interests of protecting against cyber-attacks. Like most cyber security efforts, the CMF is in the early stages of its operation and has only recently been deemed fully operational in early 2018. Thus, research on establishing training models for these new cyber teams is in development. The creation of a comprehensive training program known as the Performance Assessment Suite (PAS) is one of the more ambitious projects currently in the research stage (Buchler, Trent, Kelley, & Aved, 2017). The PAS aims to utilize observations of cyber team users in the field as well as physiological sensor data in order to determine behaviors which are amenable to successful cyber team performance as a foundation to train future CMF team members. Due to the wide scope of the project's goals, this research is broken down into four distinct thrusts.

Thrusts of the PAS

The first thrust of the research program will utilize simulated environments to examine CMF team performance under fluctuating circumstances with the aim to determine which team characteristics and practices lead to the success when combating a cyber-threat. The second thrust will appraise the validity of various psychophysiological measurement equipment for predicting subjective state and, ultimately, team performance. Specifically, this equipment should provide researchers the ability to track cognitive workload, visual attention, decision-making, and affective state in real time during cyber missions (Buchler et al, 2017). Viable equipment must be accurate and comfortable for the CMF team member as they must be able to wear it without distraction over the course of cyber tasking while simultaneously recording valid physiological data with little to no signal interruptions. Thrusts three and four combines insights from the first and second thrusts to create and validate behavioral and state models of the CMF team and its members with the aim to create a program which can track and manage team performance and effectiveness (Buchler et al, 2017).

This project evaluates the efficacy of select physiological measurement equipment within the scope of the second thrust of the PAS. The goal of this project is to determine whether an equipment type with higher ease of use can accurately track physiological changes during missions as well as another reliable equipment type that is more intrusive. This project aims to evaluate the performance of the Empatica E4 by comparing physiological data recorded from the device to another, validated physiological recording device as well as examining convergent validity with a subjective measure.

Subjective analysis of stress

Defining stress is a complicated endeavor. In the context of this project, I will be defining stress as a transaction. The transactional theory of stress, in this context, is best fit to describe the effects of individual differences in stress responses over different stressor stimuli within the scope of cyber team performance.

Transactional Theory of stress

The Transactional Theory of stress (otherwise known as the Cognitive theory of stress and coping) looks at the interaction between the person and their environment (Lazarus & Folkman, 1984). The transactional theory defines stress as the imbalance between demands and coping resources when in the presence of a threatening stimulus. The transactional theory is unique because it is a dynamic theory, which accounts for individual differences when presented with a given stressor (i.e., a stress inducing stimulus). Different people will react to the same stressor stimuli in unique ways just as a person will react to different stressors with different reactions. Thus, a person's interpretation of an event may be more important for understanding a person's stress response than is an understanding of the event itself. The transactional theory postulates that in the presence of an ambiguous stimulus, people conduct multiple appraisals in order to determine whether a stimulus is threatening, and the corresponding response.

Primary appraisal is the first stage of the appraisal process. Primary appraisal is not necessarily a conscious process and often happens automatically upon exposure to an ambiguous situation. The situation is evaluated to determine the potential risks to the person. This evaluation is highly subjective and risk analysis varies between people depending on individual factors (e.g., experience with event). The analysis compares the threat against the personal goals of the individual. The person's evaluation of these two components is imperative to appraisal. There are three potential outcomes to the primary appraisal of a situation. In neutral (irrelevant) and positive (benign – positive) appraisals the person does not believe the situation contains significant risk and a stress response is not triggered. The third outcome is a negative appraisal, which signifies an imbalance between the resources and demand of the situation.

Secondary appraisal operates in concert with primary appraisal and deals with coping methods in order to overcome the stress inducing stimuli. Coping is the process of changing actions and behaviors in response to unique stressors which overload the resources of the person (Lazarus & Folkman, 1984). In this instance, the stressor cannot be avoided, and one must instead find a way to deal with the situation. Like the original analysis of the stressor stimuli, coping methods are highly subjective and are unique to both the situation and the person. Matthieu and Ivanoff (2006) propose that there are two types of coping methods: emotion-focused and problem-focused coping methods. Emotion focused coping methods involve the person under stress processing the emotional and psychological effects of stress directly with the goal of improving his or her sense of psychological wellbeing amid stress. Emotion-focused coping has benefits in some circumstances (e.g., repeating motivational statements to oneself in the face of a task) but may also be a detriment in other forms. This is evidenced by avoidance tactics common in emotion – focused coping which avoids stress, often preventing effective behavioral response. Problem focused coping evaluates the problem at hand and seeks ways to eliminate the problem. Analysis of coping methods when in the face of task stressors by Matthews and Campbell (1998) demonstrated the individual differences in traits between those who adopted problem – centered coping methods (higher self – efficacy) and emotion-focused coping methods (increased neuroticism and trait anxiety) when given a similar mentally demanding task. The interplay between these two coping methods is important to resolve the stressful situation and is highly influenced by many psychological factors, one of which (anxiety) will be discussed in detail below.

Anxiety

Anxiety is often defined interchangeably with stress. While both are related, the origins of both differ. Stress generally responds to external causes (i.e. stressor stimuli) and is generally a temporary state of mind which arises in the face of daunting circumstances. Usually, stress is eliminated once the stressor is removed. Anxiety is a common reaction to stress. It is an emotional state which can persist after the elimination of a stressor stimuli or even before the presence of another. Anxiety is characterized by underlying feelings of tension and dread as well as physical changes such as increased blood pressure. Anxiety, in the context of the transactional theory, is influenced by the primary appraisal of a stressor stimulus and generally leads to a higher occurrence of negative appraisals. Anxiety can also be brought on as a result of a negative appraisal to a situation. The presence of anxiety is detrimental to the ability to self-regulate emotion and can increase the resources needed to cope with a stressor stimulus. These factors negatively influence performance and subsequently increase anxiety, creating a vicious cycle which can exacerbate the stress response and decrease the effectiveness of coping strategies. Anxiety as a detriment to coping with stressor stimuli is evident in performance in test taking (Matthews, Lin, & Wohleber, 2019), driving (Matthews et al, 2019) and athletic competitions (Matthews et al, 2019; Wilson, Peper, & Schmid, 2006).

State – Trait Personality Index

Measuring the effects of stress on user performance relies on analysis of subjective (independent thoughts and emotions of an individual) and physiological (focuses on the processes of the body) measures. Subjective analysis of stress involves understanding the mental state of the user. A *state* is defined as a temporary pattern of mental processes which is associated with emotional, cognitive, and motivational changes (Matthews, Szalma, Panganiban, Neubauer, & Warm, 2013). In order to obtain information on the subjective state of the user, surveys are provided. For this experiment, the State – Trait Personality Index (STPI) will be utilized to obtain subjective stress state information.

While surveys can provide some subjective insight into a cyber-team member's stress in facing a challenging cyber event, the utilization of physiological metrics such as heart rate variability (HRV) provide alternative stress assessment opportunities, which can be used without interruption of tasking and may have the potential to provide insight beyond what cyber team members are able to provide through introspection. Below I will discuss relevant physiological measures and metrics and how they can be used to detect stress during stressful events.

Physiological analysis of stress

Autonomic Nervous System function

The Autonomic Nervous System (ANS) is an important brain system utilized by researchers to discern physiological measures of stress response. ANS regulates many internal bodily functions such as heart rate and digestion (Schmidt & Thews, 1989). ANS functions are typically beyond direct conscious control (some autonomic functions like respiration, swallowing and coughing can be consciously moderated). The ANS is separated by two distinct subsystems: the Sympathetic Nervous system (SNS) and Parasympathetic Nervous System (PNS). The SNS is activated in the presence of external stimuli and internal thoughts and is linked to arousal of several of the body's systems to prepare the body for action. Physiological responses like increased heart rate, dilated pupils and increased skin conductance are indicative of this arousal as well as the occurrence of the "fight or flight" response. In contrast, the PNS is usually activated during times of relaxation and digestion and is generally active during periods in which internal body systems are slowed in order to conserve resources. The PNS usually works to induce effects to bodily systems that counter those of SNS activation (i.e., reduced heart rate and decreased skin conductance). In most instances, the PNS and SNS are complementary systems whose functions are usually depressed with the activation of the other – meaning that, for example, when the body is in a state of SNS activation, PNS activity is minimal and vice versa. It is important to note that while in most cases, activation of one system is tied to depression of the other, there are instances where both the ANS systems may be aroused or depressed (such as sexual arousal or being anesthetized) (Porges, 1995). ANS activity relates to the physiological activation of a stress response (Porges, 1995). The link between ANS activation and stress can be applied within the context of the Transactional Theory of stress. Specifically, in the presence of a significant threat (e.g., a cyber-event), if a person believes the threat to be real and consequential (primary appraisal) and believes that they may lack the resources for coping with the threat (secondary appraisal), one would expect to see ANS activity which corresponds to the user response to the stressor. In this case, a PNS dominant response to a stressor stimulus would be indicative of a person who does not identify the stressor stimulus as a threat or has adapted to the presence of the stressor and is able to cope with it. A SNS dominant response may be an indicator of the perception of the stressor stimulus as a threat and the mobilization of fight-or flight systems. ANS activity and stress response can be monitored and analyzed utilizing physiological metrics like heart rate variability (HRV).

Heart Rate Variability

Heart rate variability is regulated by the interactions between the PNS and SNS. According to Lehrer and Eddie (2013), heart rate variability (HRV) is a measure of the inter-beat interval (IBI) which can be recorded using an electrocardiogram (ECG). ECG readings derive heart rate by measuring the electrical activity of the heart during a heartbeat. The R interval is derived from the QRS complex of an ECG reading of a heartbeat, which is defined as its main ‘spike’. The R-wave or ‘R-peak’ of any given reading is representative of “the electrical timing of the heart” (Malik & Terrace, 1996). The timing between any two R – peaks in an ECG reading is irregular or ‘variable’. For example, given a sample ECG reading that states that there are 60 R-peaks in one minute, the time between each peak will not uniformly be 1 second – rather it will vary between any two individual spikes to total 60 spikes in one minute. For example, the time between two peaks in an ECG recording can be 1.05 seconds and the very next two peaks could occur in .95 seconds. the ability for heart rate to be variable is highly important for ANS processes. There are many situations where a person will need to adjust to external stimuli quickly and must shift from a state of rest to one of action. This shift will require a rapid change from increased PNS functionality to increased SNS functionality effectively without overloading the body. HRV operates in ‘negative feedback loops’ with blood pressure via the baroreflex. The negative feedback loop between HRV and blood pressure enables adaptations based on situational demands (Lehrer & Eddie, 2013). High HRV scores can be viewed as a sign of stress management and the ability to control emotional responses in the presence of stress (Appelhans & Luecken, 2006).

Low HRV is indicative of reduced parasympathetic cardiac control (Evans, Seidman, Tsao, Lung, Zeltzer, & Naliboff, 2013). HRV is also sensitive to acute stress (Evans et al., 2013). Lower HRV has been linked to higher morbidity rates in those with congenital heart diseases as well as an increased severity in the symptoms of illnesses ranging from asthma to depression (Gevirtz, 2013). Studies by Evans et al. (2013) showed the children with chronic abdominal pain demonstrated a significantly lower resting HRV measure than healthy children. In addition, children with chronic pain were shown to have an unchanging HRV response to pain, contrasting with a significant HRV response in healthy children. HRV is a marker for cardiac health by depicting the balance between the PNS and SNS (Peper, Harvey, Lin, Tylova, & Moss, 2007).

Time Domain HRV assessment

HRV can be measured utilizing time domain indices. Time domain indices examine the variability of the IBI over a specific time period and quantify the raw information (Shaffer & Ginsberg, 2017). Statistical analysis of these raw IBI markings are evaluated to determine the average variability over the defined period (Vivosense, 2017). Values derived from time domain indices are representative of autonomic activity (Vivosense, 2017). Measures like the standard deviation of the normal waves (SDNN) are influenced by both the PNS and SNS while the root mean square of consecutive normal waves (rMSSD) are indicative of PNS activity (Shaffer & Ginsberg,2017). Normal waves are defined as R waves that are not abnormal - when analyzing data, abnormal waves are eliminated, and the remaining waves are calculated for HRV indices. Time domain recordings vary in length of term from ultra-short (less than 5 minutes), short (approximately 5 minutes) to long (24 hours). Use of ultra-short recordings are limited due to problems with methodology and effectiveness of the recordings (Shaffer & Ginsberg, 2017), yet recordings lasting longer than 5 minutes have shown to be able to assess HRV accurately. Short and long term SDNN and rMSSD analyses have been employed in the past to examine the effects of HRV in cardiac health. SDNN was noted for its accuracy in predicting morbidity and mortality of cardiac issues over 24-hour terms (Kleiger, Miller, Bigger, & Moss, 1987). rMSSD scores derived from short term 1 – hour recordings of seizure victims were able to determine the probability of risk for sudden unexplained death by epilepsy (Degiorgio, Miller, Meymandi, Chin, Epps, Gordon, & Harper, 2010). In this study, SDNN and rMSSD scores will be utilized as a source to compare the data collected from two devices.

Description of Recording Equipment

Calculating heart rate variability metrics in real time requires specialized equipment capable of accurate recordings. For this project, I will be comparing the Empatica E4 and the Equivital EQ02 to determine the similarities in the recording and analysis of time domain HRV metrics.

Equivital EQ02

This project will utilize the highly specialized, medical grade Equivital EQ02 for recording HRV metrics via electrocardiograph (ECG) readings. The EQ02 is designed to fit snugly around the chest of the user in order to limit data interruptions via movement of the recording device. Experimental evaluations on the EQ02 showed that HRV data recorded from the device were both reliable and valid (Liu, Zhu, Wang, Ye, & Li, 2013). The EQ02 also compared favorably to other highly rated medical grade devices in the quality of ECG and HRV data extracted (Akintola, Pol, Bimmel, Maan, & Heemst, 2016). While the EQ02 is effective in recording and extracting viable HRV data, there are issues with putting on and taking off the EQ02. The need to have the EQ02 life shirt directly on the skin requires for users to strip their upper body in order to both put on and take off the equipment, a practice which can take a significant amount of time. Female participants may have additional issues with the EQ02, as discomfort may be present if they need assistance putting on the life vest. A potential dilemma may occur if there is no female researcher that the participant is comfortable with assisting in vest setup. It is because of these ease of use issues that another, less invasive type of recording equipment will be compared to see if the data collected are similar.

Empatica E4

The Empatica E4 is a wristband that tracks HRV metrics using a Photoplethysmogram (PPG) sensor. A PPG sensor acquires physiological information by flashing a light directly on the skin and measuring the blood volume in the veins. The E4 is designed to be placed on the user's wrist due to the presence of multiple blood vessels in the wrist (an ideal location for recording PPG information) while being noninvasive and relatively fast to put on and remove. However, PPG readings may be interrupted due to the movement of the band. Evaluations of the quality of E4 recordings have compared favorably to stationary recording devices for detecting stress in high workload conditions (Ollander, Godin, Campagne, & Charbonnier 2016). Considering that the E4 is noninvasive, an analysis to see if there is any difference in the quality of data between the E4 and Eq02 is desirable to determine which piece of technology is optimal for use by cyber team users.

PPG and ECG recording comparison.

The EQ02 and E4 are both medical grade devices that have been empirically tested with favorable results, but both utilize different sensor types to record data. This discrepancy in recording may result in a difference in the quality of readings from each device when deriving HRV metrics. Selvaraj et al. (2008) found that PPG lagged ECG when examining the time required for transmission of wave pulse yet showed insignificant differences in the quality of HRV indices between the two methods. Due to the difference in sampling rates, there may be issues with PPG readings tracking accurately with ECG readings in short term sessions (> 5 minutes) ("ECG vs PPG for Heart Rate Monitoring: Which is Best?", 2015). Other experiments examining the effectiveness of both methods have stated that the PPG can be a legitimate alternative for obtaining HRV metrics to the ECG (Lu, Yang, Taylor, & Stein 2009; Bolanos, Nazeran, & Haltiwanger, 2006), indicating that the disparity between both measurements may not be significant. For this experiment, the tradeoff between convenience of the equipment and the accuracy of the readings between the two measuring devices will be tested in order to evaluate which device is more practical in obtaining real time metrics of cyber team users.

Project Summary

As a proxy for the cyber team user in the PAS, this project is leveraging an investigation of stress management in police cadets, who took part in a simulation which emulated a high-risk situation they might encounter as an officer. This situation creates a significant stressor. Differences in stress throughout a lengthy training period were utilized as an opportunity to record subjective and physiological data. The cadets wore both the E4 and EQ02 while performing the simulation – creating an opportunity to compare the quality of data between devices during a high stress situation. The aims of this project are as follows:

- Aim 1A: Compare HRV metrics between the E4 and EQ02 during baseline. I hypothesize that despite the differences in method of recording, the E4 will not significantly differ from EQ02 device when calculating HRV in line with evidence of previous work
- Aim 1B: Compare HRV metrics between the E4 and EQ02 during final simulation. I hypothesize that despite the differences in recording methods, HRV data derived from the E4 and EQ02 will not significantly differ.
- Aim 2: Evaluate whether the E4 can accurately track changes in HRV data at baseline and during the simulation. I hypothesize that the E4 will be able to accurately track changes in HRV data.

- Aim 3: Determine whether E4 readings correlate with STPI scores. I hypothesize that the E4 will have a strong positive correlation with STPI scores based on the HRV recordings.

Method

Participants

Eight (8) participants were enlisted for this experiment. Participants were police cadets in Orange county attending the School of Public Safety at Valencia College. Participants were between the ages of 20 to 33 years old, six male participants and two females. Program requirements for Valencia called for mental health screenings as well as a physical fitness evaluation, thus eliminating participants that would otherwise have been removed for health issues. Dropping out or failing the course was ground for removal from the experiment. No vulnerable populations were used. Participants were given an identification number in order to maintain confidentiality. This number would be the only identifier for this experiment.

Procedures

Police cadets were given a presentation during the first few weeks of instruction in their classroom at the College of Public Safety at Valencia College explaining the experiment. This presentation focused on the effects of stress on performance, both physical and cognitive, and stress mitigation techniques and their benefits. The presentation then detailed the experiment and the tasking for the cadets. Interested members of the presentation were offered the opportunity to volunteer for the study after the presentation concluded.

The participants who chose to volunteer for the study were taught resonance breathing over a 12-week period. Over this 12-week period, the participants were to meet with the researchers 10 times for resonance training sessions and data recording. In the initial session, participants signed an informed consent form and were given their unique identifier. After signing the consent form, the participants were instructed to put on the EQ02 and the E4 (due to the limited availability of

E4's for the project, not every participant was given an E4 for weekly sessions). Once the equipment was put on the participants, the participants were instructed to fill out a State Trait Personality Index (STPI: Spielberger & Reheiser, 2004). The STPI can be administered quickly and easily (Spielberger & Reheiser, 2004), making it advantageous to give consistently in order to track progress over long periods of time. A 10-minute baseline in which the participants were told to sit quietly and relax was conducted afterwards to discern resting physiological data as well as to warm up the equipment.

After the baseline, participants were instructed to download the Heart Rate + Coherence application by SoftArea. This application is available for both Android and Apple devices. The researchers purchased the Heart Rate+ Classic add on (a 5-dollar purchase) for the participants. The Heart Rate+ application utilizes a guided breathing counter as well as a heart rate sensor (through the light from camera flash) in order to measure heart rate through a counter in the middle of the screen.

Once the application was installed, the researchers adjusted the breathing type in order to minimize the hold to as little as possible, shorten the inhale to about 20% of the breath time and maximize the exhale to as high as possible (depending on the model of the application, this ranged between 70% - 80%). After this change, the participants were instructed to test different breathing lengths from 8 – 12 seconds at this breathing type and to write down their maximum and average coherence scores. Once the participants recorded these scores, the researchers determined the resonance rate which the participants would train at for the entirety of the experiment by examining at which resonance rate the participant had the highest maximum resonance score. If there were multiple resonance rates with the same max score, the rate with the highest average score was selected. The participants were told that they must practice every day for 20 minutes at this selected breathing rate and to come back in for additional sessions with the researchers once a week. After each use of the EQ02 life vest and module, both were washed to ensure both were sanitized for every session.

The rest of the sessions involved the participants meeting with the researchers in order to fill out an STPI form, put on an E4 (if applicable), write down their daily session scores and to conduct a 20-minute weekly breathing session. These sessions served the purpose of collecting STPI data as well as giving the researchers an opportunity to evaluate the way the participants were breathing. The participants were given another baseline reading with the EQ02 at the 6^{th} session with the researchers, marking the midpoint of their training.

The final time the researchers met the participants, the cadets engaged in a simulation in the VirTra simulator in order to evaluate their cognitive and physical performance under stress relevant to competent law enforcement. This scenario involved a high-risk situation the cadets may encounter in the field. This scenario is used in cadet training and provided by the school. The cadets were graded on their performance by training officers.

Upon arrival, all participants put on the Life Vest and Empatica E-4 wrist band. The experimenter took baseline physiological measures and the Stress – Trait Personality Index (STPI) was taken prior to the scenario. Participants performed in the scenario one at a time. The cadet's performance was graded and recorded by the researchers and the police instructors.

Apparatus

Equivital EQ02 Life Monitor

The Equivital EQ02 Life Monitor is a lightweight sensor belt (Figure 1) with embedded respiratory function sensors woven into a sensor strap which is worn around the chest of the participant. Heart rate and respiration rate measures are provided by these sensors, both of which are used to derive heart rate variability measures.

Figure 1: Equivital EQ02 Life – Shirt (Left) and SEM Module (Right)
Source: AD Instruments product website
https://www.adinstruments.com/products/equivital-sensor-belt
https://www.adinstruments.com/products/equivital-sem-sensor-electronics-module

Empatica E4 Wristband

The E4 is a Wristband containing PPG and electrodermal activity sensors to track psycho-physiological data such as blood volume pulse (from which HRV can be derived). The E4 also includes a 3 – axis accelerometer to track movement and event mark button (see Figure 2) to tag events in the readings. The E4 includes a mobile application which allows for real time analysis of information while on the user.

Figure 2: Empatica E4

Source: Empatica website
https://www.empatica.com/research/e4/

Questionnaires and Tests

Participants were asked to complete the STPI prior to engaging in the simulation. The STPI is a 10-question, 4-point scale questionnaire, ranging from scores from 0 (not at all) to 4 (very much so). The STPI analyzes metrics of anxiety, anger, curiosity and depression to measure the present emotional states and the underlying temperament or traits of each metric on a participant (Spielberger, & Reheiser, 2004) with questions such as “I feel calm” to examine these metrics.

Design

Repeated measures two-way ANOVAs were run evaluating time domain HRV metrics (rMSSD and SDNN) derived from physiological recording equipment (EQ02 and E-4) and the stress condition (baseline and simulation) in order to evaluate the ability of each physiological device to measure differences in stress between the two conditions for HRV.

Multiple Pearson correlations were run on the HRV data (through rMSSD metrics) obtained from the E4 and the STPI scores of the participants to determine the relationship between the E4 data and the STPI scores.

Results

In order to investigate the impact of different stress conditions on the recordings gathered from both equipment types studied, 2 (EQ02 and E4) X 2(Baseline and Simulation) repeated – measures two way ANOVAs were run on two time domain data indices (rMSSD and SDNN). A correlation was then run on time domain data derived from the E4 and STPI scores obtained by the participants. For this analysis, I will report trends with a *p* – value less than .05 as significant.

Comparisons between the EQ02 and E4

rMSSD

The results for the rMSSD data were mixed. There was a main effect for equipment differences, $F(1,8) = 6.70$, $p = .041$, as well as an interaction effect between the equipment and the conditions, $F(1,8) = 8.138$, $p = .029$. There was no significant main effect for the conditions, $F(1,8) = .002$, $p = .967$. Graphical analysis of the two equipment types in both conditions show different trends for each equipment type (see Figure 3). The results of the ANOVA suggest a significant difference in the values of the rMSSD data obtained from the E4 and EQ02 – which rejects the null hypothesis for aims 1A and 1B. Due to the changes in the rMSSD scores detected by the E4 in the baseline and simulation conditions as shown in figure 3, the null hypothesis for aim 2 could also be rejected. It is important to note, however, that the trends for the E4 and EQ02 of rMSSD scores were the opposite of each other, suggesting that in terms of E4, it may track differences in HRV metrics, but it may not track them correctly.

SDNN

There was no main effect for equipment type, $F(1,8) = 4.65$, $p = .074$ or for condition, $F(1,8) = 1.55$, $p = .260$. There was also no significant interaction between the equipment types and conditions, $F(1,8) = 0.90$, $p = .778$. Unlike in the rMSSD measures, SDNN scores for both equipment types followed a similar trend between conditions (see Figure 4). The results of this analysis suggest that SDNN data derived from the E4 and EQ02 were not found to be significantly different – a failure of rejection of the null hypothesis for Aims 1A and 1B. It must be noted that although there are some similarities in the data, the *p*-value for the main effect was only slightly over the .05 cut-off point, suggesting a weak similarity.

Correlational Analysis

In order to evaluate the final aim, Pearson Correlations were run on rMSSD data extracted from the E4 and STPI results for participant sessions with at least 5 data values. Of the 9 sessions which participants were administered the STPI and given E4s to record HRV data, 4 had enough participants to run a correlation. The first correlation (second session) did not show a significant correlation between rMSSD data and STPI scores, $r(5)= .039$, $p = .950$ (see Figure 5). The next correlation was run on the sixth session and the correlation was not statistically significant, $r(5)= -.304$, $p = .619$ (see Figure 6). A correlation run for the eighth session was not statistically significant, $r(5)= -.794$, $p = .109$ (see Figure 7).The final correlation was run during the final Vitra simulation and was not statistically significant, $r(6)= -.336$, $p = .516$ (see Figure 8). The correlation results suggest that rMSSD data does not have any significant correlation to the results of the STPI scores – failing to reject the null hypothesis.

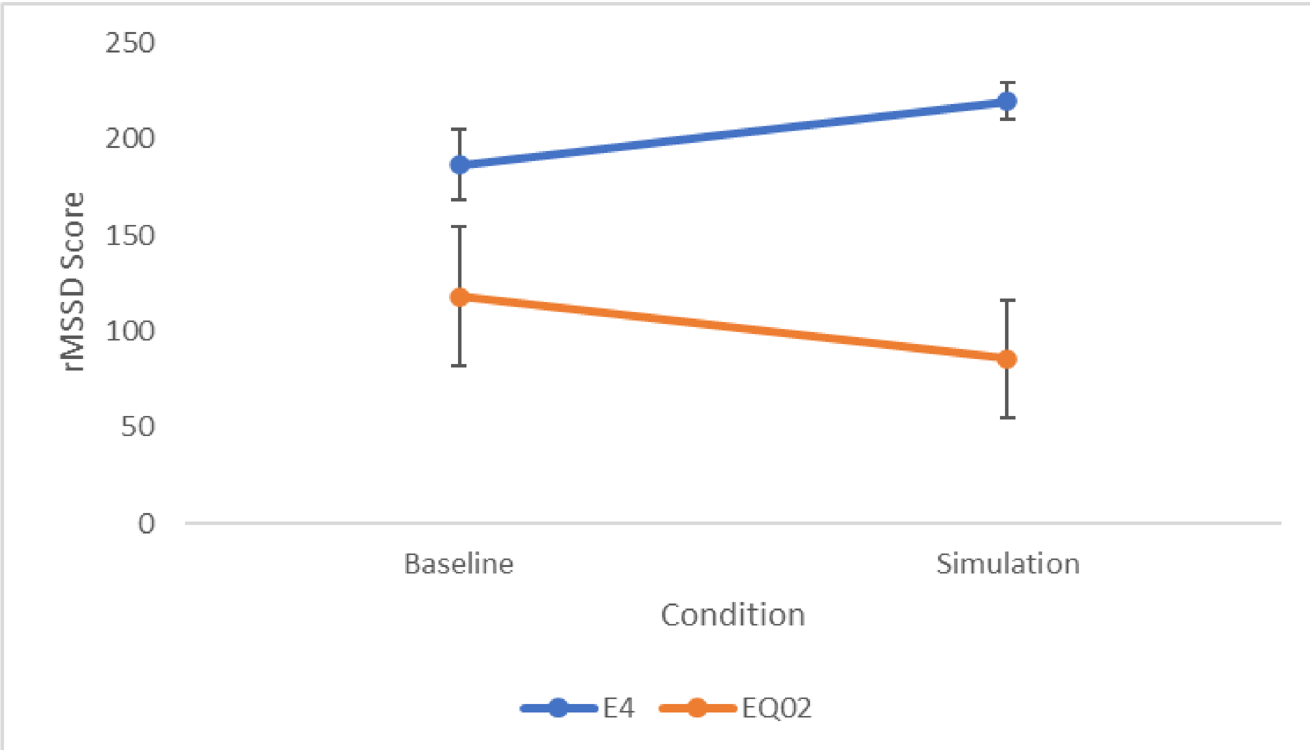

Figure 3: Average rMSSD scores for E4 (1) and EQ02 (2) over baseline (1) and simulation (2) conditions

.

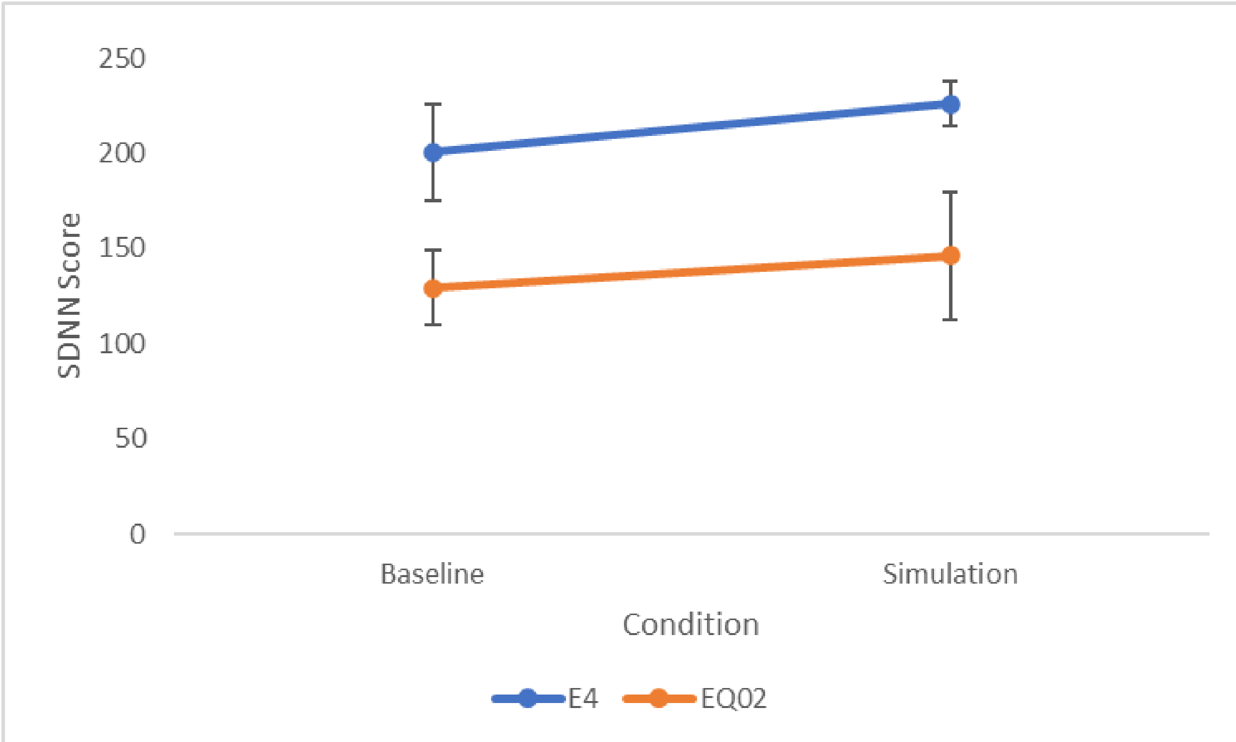

Figure 4: SDNN trends between E4 (1) and EQ02 (2) over baseline (1) and simulation (2) conditions.

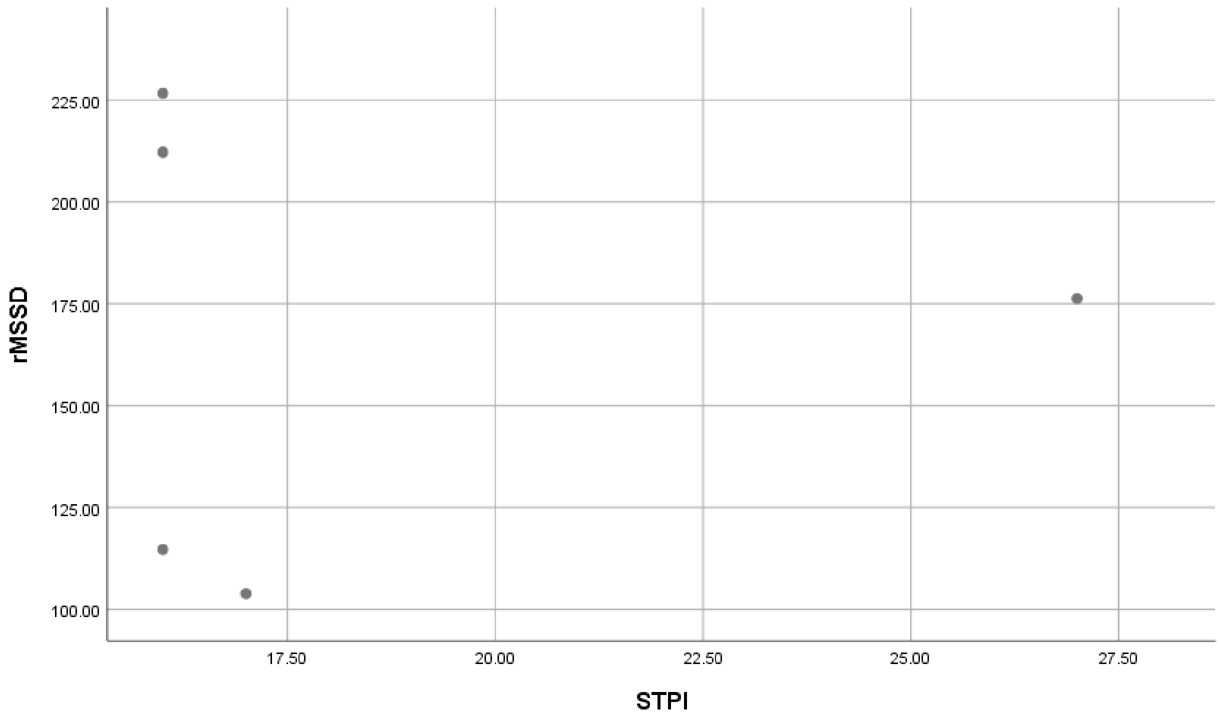

Figure 5: Correlation between STPI scores and rMSSD scores derived from the E4 (second session)

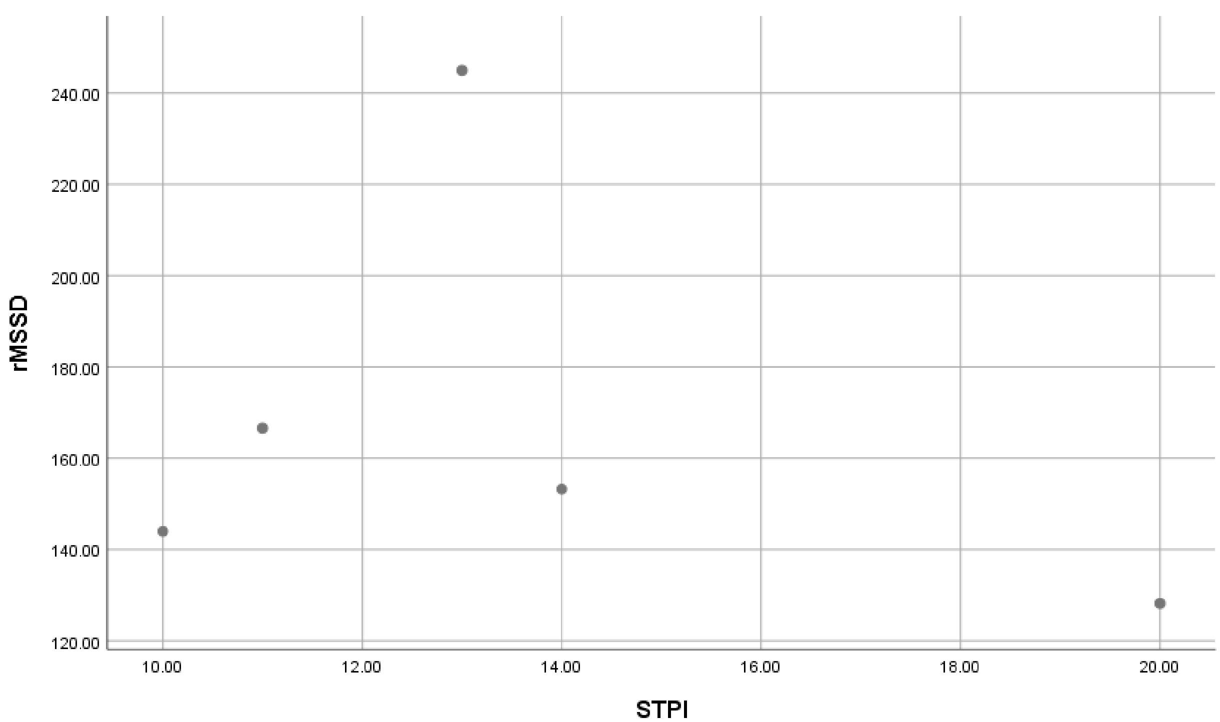

Figure 6: Correlation between STPI scores and rMSSD scores derived from the E4 (sixth session)

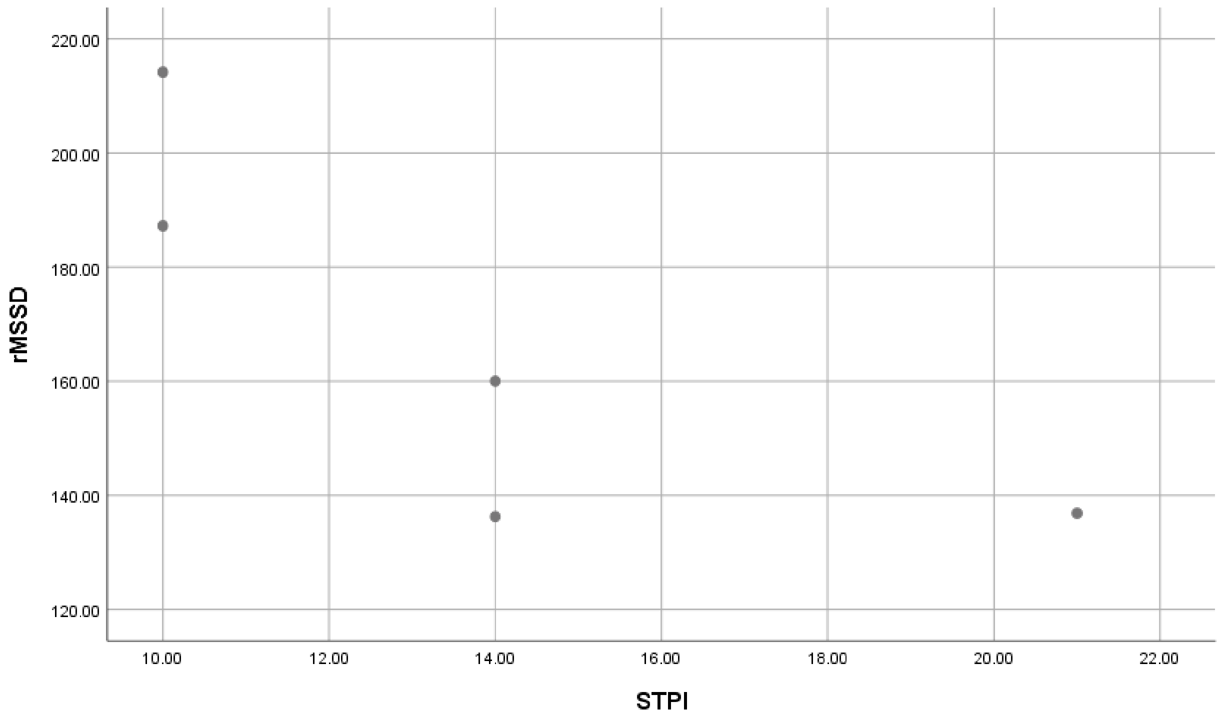

Figure 7: Correlation between STPI scores and rMSSD scores derived from the E4 (eighth session)

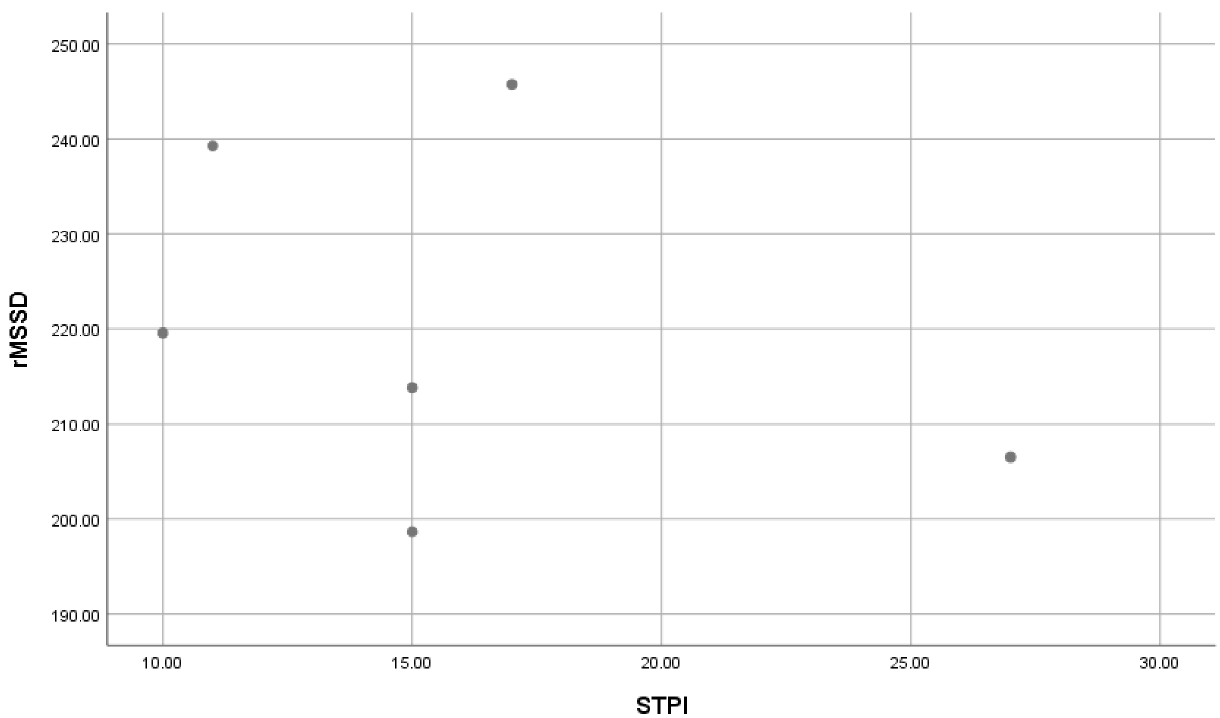

Figure 8: Correlation between STPI scores and rMSSD scores derived from the E4 (simulation)

Discussion

This study examined the viability of HRV time domain data extracted from the Empatica E4 PPG wristband by comparing data with another physiological recording device, the Equivital EQ02. The ability for the E4 to accurately record differences in stress conditions was also tested. This study also aimed to find convergent validity between the E4 time domain data and subjective results gathered from STPIs collected over researcher guided sessions with police cadets. The results showed that there was a significant difference between the E4 and EQ02 for one metric and that there was no convergence between the rMSSD scores derived from the E4 and STPI data.

Aims 1A and 1B examined time domain HRV data derived from the Empatica E4 and Equivital EQ02 in order to determine whether the information gathered was comparable. Utilizing two distinct time domain metrics (rMSSD and SDNN), results determined that there was a significant difference in rMSSD data collected, signifying large differences in the values obtained between devices and a rejection of the hypothesis that the devices would be able to track HRV metrics somewhat similarly. Looking at the figures, this conclusion is evident in the trend lines of the Equipment charts, which show that as conditions changed, rMSSD in the E4 would decrease, while rMSSD in the EQ02 would increase.

Conversely, SDNN data extracted from the E4 and EQ02 were found to not be significantly different when analyzed. This suggests that data between the E4 and EQ02 were similar enough to fail to be significantly different from each other, affirming the original hypothesis. It should be noted, as was stated in the results, that the p – value for the main effect was .074, only slightly above the threshold to be significantly different.

One possible explanation for these results is the differences in the percentage of artifacts (erroneous heart beats which are eliminated before processing data) in the data sets of each device and its effect in analyzing different metrics of HRV. A high number of artifacts in a data set has the potential to pollute the data set by creating unwanted detections of normal waves in the data and make it unlikely to measure each heartbeat accurately (Vivosense, 2017). Heartbeat detection errors, even with a percentage error rate as low as 2%, have the potential to create calculation errors when determining HRV metrics (Vivosense, 2017). These errors may have been present in the analysis of the time domain metrics, specifically the rMSSD analysis. The rMSSD is a very specific metric utilized to examine PNS activity in a heart rate reading. This means that in order to obtain reliable readings, a significant amount of data must be normal waves in order to accurately track PNS activity over a given time period. A high number of artifacts can confound the information in the data set and misrepresent the levels of PNS activity in the data set. Broader metrics like SDNN look at the overall activity of the ANS including the SNS so it is much more resistant to the presence of artifacts in a data set.

There are many forms of artifacts, but the most common type is the motion artifact, which is caused by a user movement which is significant enough to create the appearance of an irregular heartbeat. Other artifact types include temperature changes, muscle contractions and vocalizations (Vivosense, 2017). In the case of this experiment, the location of each device on the participants body may have created a significant difference in the number of artifacts that each device detected. The E4 is placed on the participant's wrist, a location that is more likely to be affected by erroneous movement like bumping on a table or subconscious movements (such as moving the arm to another position to increase comfort) then the EQ02's positioning on the chest. I believe that the higher frequency of movement artifacts in the E4 greatly affected the quality of data gathered, especially when considering the E4's sampling rate.

The sampling rate of a device determines how many times in a one-minute span the device records a data point. When analyzing an HRV data set, the sampling rate (measured in Hz) of the equipment is important when considering the number of artifacts within that data set. If a data set recorded from a device with a low sampling rate has a high number of artifacts, the number of normal waves available for data analysis of HRV metrics will be low, and the results may be polluted and a possibly unusable for the sake of drawing conclusions from the data. For this experiment, the EQ02 ran at a sampling rate of 256 Hz while the E4 ran at only 64 Hz. For sensitive time domain measures like rMSSD, a lower sampling rate could be particularly worrisome if there is a high risk of movement artifacts (such as in this experiment). If there is a low number of data points being analyzed from the outset and a significant percentage of them are not normal waves, the ability for the rMSSD in the E4 to accurately trend PNS activity becomes compromised. As was mentioned before, metrics like the SDNN are much more resistant to artifacts due to the analysis of overall ANS activity in HRV calculations rather than an aspect of the readings. The importance of sampling rates in time domain HRV data has been tested in the past with similar conclusions. In an analysis of the capabilities of different sampling rates recording fetal heart rate variability of sheep to detect increased toxicity in the blood (acidemia), Durosier, Green, Batkin, Seely, and Ross (2014) determined that not only did the lower sampling rate not detect any significant differences in HRV between baseline and experimental conditions, but it was also unable to detect fetal acidemia in each of the 9 fetuses examined for the experiment.. This discrepancy in sampling rates combined with the increased likelihood of movement artifacts of the E4 is a possible explanation for the mixed results in data similarities between time domain indices between devices.

Another possible limitation of the E4 which may have affected the time domain data needed for Aims 1A and 1B is the E4's inconsistency in staying connected to the E4 manager application. For a session to record properly, the wristband must always be connected to the application. During the training sessions with the researchers, E4s would occasionally disconnect from the managers which caused the data set to be incomplete and unusable for analysis. This error happened on multiple occasions, including during the final simulation, and is a cause for concern for those who wish to use the E4 or devices like the E4 in the future.

Aim 2 looked at the ability of the E4 to track changes in HRV metrics between conditions at baseline and at the simulation. For rMSSD measures, the E4 was able to track changes in HRV metrics from one condition to the other. The E4 was not able to effectively track changes in SDNN measures, however. What is interesting is the mixed results in its recording when compared to the EQ02. The E4 was able to track changes in rMSSD scores but the trend was not the same as the one from the EQ02. While the EQ02 demonstrated higher rMSSD scores from the baseline to the simulation, the E4 showed lower rMSSD scores from the baseline to simulation. This relationship is concerning and can be attributed to the high numbers of artifacts in the data set polluting the results. SDNN scores recorded from the E4 tracked in a similar fashion to the EQ02.

Aim 3 attempted to see if there was any convergent validity between rMSSD data derived from the E4 and STPI survey data obtained via the participant sessions with the researchers. Pearson correlations run on four participant sessions found that there were no significant correlations between the rMSSD data and the STPI scores, rejecting the hypothesis that the two would have a significant comparison. These results could have been influenced by response bias on the STPI. It is possible that due to participant knowledge that they should be relaxed when participating in the sessions with the researchers, they filled out the surveys with the belief that they should be relaxed, manipulating their scores. It is possible that weekly administration of the same survey made the participants annoyed and they filled the survey out randomly in order to complete it as fast as possible. Another potential explanation for the results of Aim 3 is the timing of the administration of the STPIs. The STPIs were given to the cadets during lunch before their daily physical training session. To many of the cadets, the physical training was the most stressful part of their day and administering the STPI before then could reflect a higher stress state than if administered after the physical training. It is also worth mentioning that due to the low sample sizes of the participants in the correlations, the power of the correlations is low.

Limitations

Another limitation to this study is one that was mentioned earlier, which was the low sampling rate of the E4. The low sampling rate of the E4 (when in conjunction with the number of artifacts) made it difficult to obtain viable HRV data from frequency domain data measures. Frequency domain measure analyze the amount of activity or "power" of ANS regions based on the frequency level. Frequency domain data would have been valuable to analyze the amount of PNS or SNS dominance occurring at each condition. The lack of normal waves in an E4 data set made this difficult to consistently obtain, so time – domain metrics were utilized instead.

The lack of experimental control was also a limitation of this study. There were multiple times within this study where the schedule of the police academy ran in conflict with the ability of the researchers to conduct training sessions with the participants in a timely manner. Instances where participants had a shorter lunch period, or an upcoming examination made it difficult for the participants to effectively fill out surveys or be available for 20-minute intervals for breathing and data collection. This inconsistency in data collection made creating a consistent schedule with the participants difficult, creating issues with communication and some frustration within the participants as any issues with the devices or applications could not be fully addressed due to time constraints.

Future Research and Conclusion

Future research with the E4 should focus on comparing HRV obtained with the E4 with other PPG sensors that are worn on the wrist. This type of research would be conducted to see if there are any overarching problems with this recording hardware that is vulnerable to movement artifacts on the wrist. This research would also serve the purpose of determining if the E4 is able to compare favorably or exceed the quality of data from similar devices.

Another avenue for future research would be the analysis of ECG and PPG devices with similar sampling rates on HRV metrics with changing stress conditions. By making the sampling rates equal, it would be a direct examination of the ability for the sensor type to extract quality HRV data. These stress conditions could take the form of a police simulation such as the one conducted in this study but it could also be utilized for any environment in which one goes from a state of relaxation to action such as taking an exam or at a competitive sporting event. This type of research would be important to see which type of sensor is the most reliable at different sampling rates.

The above evidence suggests that due to the reasons listed above, the E4 is not a viable substitution for recording HRV metrics in different stress conditions with respect to the EQ02. The issues in collecting usable data consistently combined with the increased frequency of artifacts and lower sampling rate may be enough to negate the benefits in ease of putting on and taking off the E4. More research in other PPG wrist sensors to see if there is an alternative to the Empatica E4, as the E4 proved to be too unreliable to be utilized by researchers trying to analyze physiological metrics on cyber team users.

APPENDIX A

STATE TRAIT PERSONALITY INDEX

Self-Analysis Questionnaire

STPI Form Y-1

ID number _____

Directions: A number of statements that people have used to describe themselves are given below. Read each statement and then circle the appropriate value 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 best describe your present feelings.

	Not at all	Somewhat	Moderately So	Very Much So
1. I feel calm	1	2	3	4
2. I am tense	1	2	3	4
3. I feel at ease	1	2	3	4’
4. I feel I am presently worrying over possible misfortunes	1	2	3	4
5. I feel nervous	1	2	3	4
6. I am jittery	1	2	3	4
7. I am relaxed	1	2	3	4
8. I feel worried	1	2	3	4
9. I feel steady	1	2	3	4
10. I feel frightened	1	2	3	4

References

Akintola, A. A., Pol, V. V. D., Bimmel, D., Maan, A. C., & Heemst, D. V. (2016). comparative analysis of the equivital EQ02 Life monitor with holter ambulatory ECG device for continuous measurement of ECG, heart rate, and heart rate variability: a validation study for precision and accuracy. *Frontiers in Physiology*, *7*. doi: 10.3389/fphys.2016.00391

Appelhans, B. M., & Luecken, L. J. (2006). heart rate variability as an index of regulated Emotional Responding. *Review of General Psychology*, *10*(3), 229–240. doi: 10.1037/1089-2680.10.3.229

Bolanos, M., Nazeran, H., & Haltiwanger, E. (2006). Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals.*2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. doi: 10.1109/iembs.2006.260607

Buchler, N., Trent, S., Kelley, T., & Aved, A. *Performance Assessment Suite for the Cyber Mission Force*, Performance Assessment Suite for the Cyber Mission Force (2017).

Degiorgio, C. M., Miller, P., Meymandi, S., Chin, A., Epps, J., Gordon, S., … Harper, R. M. (2010). RMSSD, a measure of vagus-mediated heart rate variability, is associated with risk factors for SUDEP: The SUDEP-7 Inventory. *Epilepsy & Behavior*, *19*(1), 78–81. doi: 10.1016/j.yebeh.2010.06.011

Durosier, L. D., Green, G., Batkin, I., Seely, A. J., Ross, M. G., Richardson, B. S., & Frasch, M. G. (2014). Sampling rate of heart rate variability impacts the ability to detect acidemia in ovine fetuses near-term. *Frontiers in Pediatrics*, 2, 38. https://doi.org/10.3389/fped.2014.00038

Eckberg, D. L, & Sleight, P. *Human Baroreflexes in Health and Disease*. Oxford: Clarendon Press; 1992.

ECG vs PPG for Heart Rate Monitoring: Which is Best? (2015, January 28). Retrieved from http://neurosky.com/2015/01/ecg-vs-ppg-for-heart-rate-monitoring-which-is-best/

Evans, S., Seidman, L. C., Tsao, J., Lung, C., Zeltzer, L., & Naliboff, B. (2013). Heart rate variability as a biomarker for autonomic nervous system response differences between children with chronic pain and healthy control children. *Journal of Pain Research*, 449. doi: 10.2147/jpr.s43849

Gevirtz, R. (2013). The Promise of heart rate variability biofeedback: evidence-based applications. *Biofeedback*, *41*(3), 110–120. doi: 10.5298/1081-5937-41.3.01

Kleiger RE, Miller JP, Bigger JT, Jr, Moss AJ. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am J Cardiol* (1987). 59:256–62.10.1016/0002-9149(87)90795-8

Lazarus RS, Folkman S. *Stress, appraisal, and coping*. New York: Springer; 1984.

Lehrer, P., & Eddie, D. (2013). Dynamic processes in regulation and some implications for biofeedback and biobehavioral interventions. *Applied Psychophysiology and Biofeedback*, *38*(2), 143–155. doi: 10.1007/s10484-013-9217-6

Lehrer, P. M., & Gevirtz, R. (2014). Heart rate variability biofeedback: how and why does it work? *Frontiers in Psychology, 5*. doi: 10.3389/fpsyg.2014.00756

Lehrer, P. M., Vaschillo, E., Vaschillo, B., Lu, S.-E., Eckberg, D. L., Edelberg, R. … Hamer, R. M. (2003). Heart rate variability biofeedback increases baroreflex gain and peak expiratory flow. *Psychosomatic Medicine*, *65*(5), 796–805. doi: 10.1097/01.psy.0000089200.81962.19

Liu, Y., Zhu, S. H., Wang, G. H., Ye, F., & Li, P. Z. (2013). Validity and reliability of multiparameter physiological measurements recorded by the equivital life monitor during activities of various intensities. *Journal of Occupational and Environmental Hygiene*, *10*(2), 78–85. doi: 10.1080/15459624.2012.747404

Lu, G., Yang, F., Taylor, J. A., & Stein, J. F. (2009). A comparison of photoplethysmography and ECG recording to analyze heart rate variability in healthy subjects. *Journal of Medical Engineering & Technology*, 1–8. doi: 10.1080/03091900903150998

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. doi: 10.1093/oxfordjournals.eurheartj.a014868

Matthews, G., & Campbell, S. E. (1998). Task-induced stress and individual differences in coping. *In Proceedings of the Human Factors and Ergonomic Society 42nd Annual Meeting* (pp. 821–825). Santa Monica, CA: Human Factors and Ergonomics Society.

Matthews, G., Wohleber, R. W., & Lin, J. (2019). Stress, skilled performance, and expertise. *The Oxford Handbook of Expertise*, 489–524. doi: 10.1093/oxfordhb/9780198795872.013.22

Matthieu, M. M., & Ivanoff, A. (2006). Using stress, appraisal, and coping theories in clinical practice: assessments of coping strategies after disasters. *Brief Treatment and Crisis Intervention*, *6*(4), 337–348. doi: 10.1093/brief-treatment/mhl009

Parnandi, A., & Gutierrez-Osuna, R. (2017). Physiological modalities for relaxation skill transfer in biofeedback games. *IEEE Journal of Biomedical and Health Informatics*, *21*(2), 361–371. doi: 10.1109/jbhi.2015.2511665

Peper, E., Harvey, R., Lin I.M., & Tylova, H. Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony? *Biofeedback* 2007; 35(2)

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

Ollander, S., Godin, C., Campagne, A., & Charbonnier, S. (2016). A comparison of wearable and stationary sensors for stress detection. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. doi: 10.1109/smc.2016.7844917

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, *5*. doi: 10.3389/fpubh.2017.00258

Schmidt, A., & Thews, G. (1989). Autonomic nervous system. In *Human Physiology* (2nd ed., pp. 333–370).

Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. K., & Anand, S. (2008). Assessment of heart rate variability derived from fingertip photoplethysmography as compared to electrocardiography. *Journal of Medical Engineering & Technology*, *32*(6), 479–484. doi: 10.1080/03091900701781317

Spielberger, C. D., & Reheiser, E. C. (2004). Measuring anxiety, anger, depression, and curiosity as emotional states and personality traits with the STAI, STAXI and STPI. In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment, Vol. 2. Personality assessment (p. 70–86)*. John Wiley & Sons Inc.

Vivosense Inc. (2017). *Vivosense HRV Professional Edition: User guide.* Retrieved from https://www.vivosense.com/hubfs/documents/heart-rate-variability-professional-edition.pdf?hsLang=en

Wilson, V. E., Peper, E., & Schmid, A. (2006). Training strategies for concentration. *Applied Sports Psychology: Personal Growth to Peak Performance*, *5,* 404–422.