Assessing Fatigue in the Field: Towards the Objective, Efficient, and Economically Viable Assessment of Acute Fatigue in On-Shift Physicians

By

Harvey Howse
Bachelor of Science, Dalhousie University, 2015

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

In the School of Exercise Science, Physical & Health Education

SUPERVISORY COMMITTEE

Assessing Fatigue in the Field: Towards the Objective, Efficient, and Economically Viable Assessment of Acute Fatigue in On-Shift Physicians

by

Harvey Howse
Bachelor of Science, Dalhousie University, 2015

Supervisory Committee

Olav Krigolson, School of Exercise Science, Physical & Health Education
Co-Supervisor

Bruce Wright, Division of Medical Sciences
Co-Supervisor

ABSTRACT

Medical mistakes made during the fatigue state result in the spread of infection, diagnostic error, psychological distress, poor patient outcomes, and ultimately, loss of life. Alarmingly, the fatigue-management systems put forth by government agency have failed to reduce the risks of fatigue in physicians. A shift from "one size fits all" approaches for fatigue management, to individualized fatigue assessment and training, is required. To date, no validated measures of fatigue are feasible for use as portable, on-site assessments. Here, I propose the use new portable EEG technologies recently validated for the collection of ERP data, as a basis for a portable fatigue assessment that is cost effective, portable, and efficient enough to be used in medical professionals. Over the course of three experiments I have provided data to support the use of the MUSE portable EEG headband, in combination with short oddball task to assess fatigue related neural impacts. Results of these experiments indicate that the P300 component is reduced in fatigued subjects in comparison to non-fatigued, and further that there is a strong correlation between subjective fatigue severity and P300 amplitude.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author would like to acknowledge funding from the Canadian Institute of Health Research, and to thank the supervisory committee and the Neureconomics Laboratory for continued support throughout the development of this project.

CHAPTER ONE: INTRODUCTION AND REVIEW

1.1 An Overview

There is no doubt that on-the-job fatigue results in employee impairment that is dangerous, and potentially fatal (Williamson et al., 2011; Dawson & McCulloch, 2005; Ackerman, 2001; Akerstedt, 2000; The Parliament of the Commonwealth of Australia, 2000; Folkard, 1997). Presently, government policies across Europe, Canada, and the US, have failed to develop effective ways to reduce the risks associated with fatigue in the workplace (Rhodes & Gil, 2002; Coplen & Sussman, 2001; Standards Australia, 2001; Mahon & Cross, 1999, Institutes BC, 1999; Gander, Waite, McKay, Seal & Miller, 1998). Within the medical context, professionals are at high risk for experiencing workplace fatigue thereby creating serious safety concerns for both patients and staff including needle prick accidents, spread of infection, diagnostic errors, psychological trauma, and ultimately, loss of life. Yet, schedule-based fatigue management systems—the most common method of fatigue management in the medical field—have been ineffective in reducing fatigue-related errors over the last decade (Drolet, Sangisetty, Tracy, & Cioffi, 2013; Antiel, Reed, Van Arenodonk, Wightman, Hall, & Porterfield, 2013; Vlopp, Rosen, & Rosenbaum, 2007; Philbert, Nasca, Brigham, & Shapiro, 2007).

Moving forward risk-management systems for fatigue must be directed away from "one-size-fits-all" regulations and towards assessment and monitoring the individual (Canadian

Medical Association, 2014). Unfortunately, just as the efforts of regulating bodies to improve fatigue-related safety outcomes in high-fatigue sectors have been lacking, the scientific community has also struggled to develop the tools necessary for improvements in the safety of these occupational sectors (Lal & Craig, 2007). Due to the complex nature of fatigue, as well as confounds of individuals' expectations, subjective psychological assessments of fatigue are insufficient in predicting safety outcomes (Johnson & Reece, 2015; Aidman, Chadunow;; Baranski, 2007; Nordbakke & Sagber, 2007; Belz, Robinson, & Casali, 2004; Ramsay, Horne & Baulk, 2004; Phillip et al., 2003; 2000; Phillip et al., 1997; Lisper, Laurell, & Van Loon, 1986; Lenne, Trigs, & Redman, 1997). Indeed, many assessments used to measure fatigue rely on questions about global fatigue or sleepiness that do not capture acute changes in fatigue and alertness (Schmidt et al., 2009).

While more promising, neurobiological and physiological measures of fatigue such as heart rate and blink patterns have also fallen short of being feasible in field use, outside of the laboratory (Aidman et al., 2015; Lal & Craig, 2007). Currently, the most potential for a neurophysiological fatigue assessment lies in a particular event-related-potential component, the P300 (Lal & Craig, 2005; See Polich & Herbst, 2000 for component). The P300 component is assessed using electroencephalographic methods and has been shown to predict task duration, subjective fatigue, and performance in both cognitive and motor tasks (Zhao, Liu, & Zheng,

2012; Schmidt et al., 2009; Gruoping & Kan, 2009; Murata et al., 2005 Uetake & Murata, 2000). Unfortunately, the nature of EEG systems limits the feasibility of using such methods as a field-based assessment beyond the laboratory (Krigolson, Williams, Norton, Hassal, & Colino, 2016; Zhao, Liu, & Zheng, 2012 Lal & Craig, 2007).

With the present study I aim to use new portable EEG technology to objectively assess fatigue on-site in professional medical environments. Krigolson and colleagues at the University of Victoria have recently validated a portable EEG headband (MUSE; IneraXon, Toronto, Ontario) for ERP-based research (Krigolson et al., 2016). Using the MUSE-EEG system I will collect P300 data from participants in a variety of field-based environments in order to explore both the relationship between the P300 and fatigue, as well as the feasibility of this system as a portable device to be used on-site in medical environments. The series of experiments proposed in this study will range in design from replication of previous ERP results to the first in-hospital pilot experiment, assessing fatigue in medical residents during a 12-hour simulation of emergency room work. In sum, this study will be the first step towards developing a consumer grade, objective, and electroencephalographic fatigue assessment usable on-site in the medical environment.

1.2 Defining Fatigue

Due to the implications of fatigue on safety, wellbeing, and health, this construct is

widely studied across multiple fields. Yet, there is little consensus as to an objective definition of fatigue. Authors in the field report that boredom, sleepiness, drowsiness, and both physical and mental tiredness contribute to fatigue, and furthermore that fatigue is also confounded by these same factors (Aidman, Chadunow, Johnson, & Reece, 2015). Indeed, even the earliest studies seeking to quantify fatigue reported on the complexity of this mental state, explaining that it relates to a decrease in a myriad of cognitive functions including alertness, memory, mental performance, and efficiency (Grandjean, 1979). Further, early research identified that these fatigue-related deficits develop both gradually and cumulatively, and that individuals may not become aware of these deficits until they reach severity (Grandjean, 1979; 1988). In this thesis work I have defined fatigue as acute impairment of cognitive function due to sustained mental effort.

1.3 Fatigue as a Safety Risk

Past literature suggests individuals' fatigue can be detrimental to workplace efficiency and safety, potentially causing harm to themselves and others (Williamson et al., 2011; Dawson & McCulloch, 2005; Ackerman, 2001; Akerstedt, 2000; Folkard, 1997). Transportation research has indicated diver fatigue as a major contributor to road accidents (Haraldsson & Akerstedt 2001; Maycock, 1997; Horne & Reyner, 1995a). The Australian government reported that fatigue accounts for up to 30% of vehicle crashes (The Parliament of the Commonwealth of

Australia, 2000). In Europe, fatigue may account for 40% of vehicle crashes (Idogawa, 1991). Alarmingly, fatigue-related accidents result in fatality rates comparable to impaired driving (Pack et al., 1995). In sum, fatigue contributes to the economic strain of individuals, industry, and community. Further, it results in injury and loss of life through vehicle crashes.

The danger of fatigue is not limited to the transportation industry. Research has linked fatigue to impairments in decision-making (Leiberman, Tharion, Shukitt-Hale, Speckman, & Tulley, 2002), language and math skills (Majekodunmi & Landrigan, 2012), working memory, judgement, recall, and executive control (Trejo et al., 2007). Indeed, fatigue results in reduced performance in cognitive and motor tasks (Eddy 2005; Beaumont, Batehat, Pierard, Coste, Doireau, & Van Beers, 2001; Harrison & Horne, 2000; Dinges, Pack, Williams, Gillen, Powell, & Ott, 1997), and is associated with an increase in error making (Gander, Merry, Millar, & Weller, 2000; Neri, Shappell, & DeJohn, 1992).

For example, when fatigue is experimentally increased by the use of monotonous tasks, or long task duration, performance on both simulated, and in-vehicle driving tasks, is reduced (May & Baldwin, 2009; Johns, Tucker, Chapman, Crowley, & Michael, 2007). Further literature compares the dangerous effects of fatigue to alcohol intoxication (Dawson & Reid, 1997). For example, Williamson and Freyer (2007) have reported that even 17-24 hours without sleep can result in cognitive deficits similar to those caused by alcohol intoxication. Indeed, inline with the

effects of intoxication, fatigue-related reductions in driving performance directly relate to an increased risk of road accidents, and thus lead to greater risks to the safety of the fatigued driver as well as other motorists and pedestrians using the roadway (Adell, Varhelyi, & Fontanta, 2011).

One of the greatest dangers associated with fatigue is the disbelief individuals have over its impact on their abilities and performance (Schmidt et al., 2009; Lal & Craig, 2007). Indeed, multiple studies report that individuals perceptions' of their performance, reaction times, and cognitive abilities during fatigue are often inaccurate (Schmidt, Schrauf, Simon, Fritzsche, Buchner, & Kincses, 2009; Moller, Kayumov, Bulmash, Nhan, & Shapiro, 2006; Belz, Robinson, & Casali, 2004; Phillip, et al., 2003; Lenne, Triggs, & Redman, 1997; Phillp, et al., 1997) This has proven to be a major issue in the area of occupational health and safety (Rhodes & Gil, 2002; Standards Australia, 2001; Baker, 2000; Institutes BC, 1999; Mahon & Cross, 1999; Gander, Waite, McKay, Seal, & Millar, 1998; *Occupational Health and Safety Act, in R. S. O. 1990; Occupational Health and Safety Management Systems (OHSAS): 18001, Canada*).

In cases in which individuals are highly practiced at a routine action, as they would be for their profession, the effects of fatigue are often severe before they are acknowledged (Canadian Medical Association, 2014). Even low levels of fatigue can affect performance, especially in situations in which circumstances unexpectedly change from the routine (Mascord & Heath,

1992). Examples of such situations would include medical emergencies managed by novice hospital staff, as well as a changing road conditions, monitored by professional drivers (Williamson, Feyer, & Friswell, 1996).

Not surprisingly, sleep deprivation research has also reported quantified performance deficits exhibited during fatigue (Nordbakke & Sagberg, 2007; Baranski, 2007; Horne & Baulk, 2004; Lisper, Laurell, & Van Loon, 1986). Yet, individuals do not perceive that their performance has changed following sleep deprivation, irrespective of their subjective feelings of fatigue (Phillip et al., 2003). Additional studies have confirmed that individuals' assessment of their performance is impaired during fatigue. In an ecological, long-duration driving experiment completed by Schmidt and colleagues (2009), participants experienced a significant drop in their accuracy at assessing their own performance during the final quarter of the task. Importantly, participants reported that in the final leg of the drive, they felt less fatigued, and more alert than the previous block of the experiment. Additionally, they believed their performance also improved at this time. Despite these perceptions the group in fact performed their worst in this block of the task. Further, all physiological measures recorded (heart rate, pupillary response, EEG signal) supported a state of reduced alertness rather than increased alertness in comparison to the prior three blocks of the experiment. Evidently, the dangers of fatigue are rooted not only in the deficits to cognitive and motor performance, but also in the refusal, or inability, of

individuals to identify and recognize their own fatigue and it's the magnitude of its effects.

1.4 Fatigue in the Medical Workplace

As introduced in the previous section, a large body of literature reports that fatigue increases the risk to personal safety in already high-risk occupational areas. Given the responsibility of medical professionals, as well as the lengthy and unstable nature of work schedules in this field (Pattani, Wi, & Dhalla, 2014), it would follow that physician fatigue is a significant risk factor in the medical environment. Indeed, physician fatigue is associated with impaired language and math skills, poor judgement, impaired decision-making, diagnostic error, and in particular, increased errors made in cases of intensive care (Eddy, 2005; Eastridge et al., 2003; Gaba & Howard, 2002). The link between fatigue and medical error is further supported by a 36% increase in medical errors occurring during overnight and on-call shifts, when physicians are more likely to experience fatigue (Aidman et al., 2015;) in comparison to during regularly scheduled shifts Grantcharoy, Bardram, Funch-Jensen, & Rosenburn, 2002). Additionally, Physicians report making more clinical mistakes when fatigued compared to being well rested (Landrigan et al., 2004; Folkman et al., 1991). Thus, there is an abundance of literature supporting a direct effect of physician fatigue on patient outcomes (Rothschild, Koehane, & Rogers, 2009; Lockley, Barger, Ayas, Rothschild, & Czeisler, 2007; Barger, Ayas, Cade, Cronin, & Rosner, 2006; Arnedt, Owens, & Crouch, 2005; Philbert, 2005; Van Dongen,

Baynard, & Maislin, 2004; Landrigan, Rothschild, & Cronin, 2004; Howard, Gaba, & Smoth, 2003; Eastrige, Hamilton, O'keef, Rege, & Valentine, 2003; Taffinder, McManus, Hul, Russell, & Darzi, 1998).

Given the above evidence, it is not surprising that between 5-20% of patients experience an adverse event regarding their safety as a patient when admitted to hospital and that as many as 50% of patients admitted are affected by some form of physician error (Baker et al., 2004, Brennan et al., 1991). The Canadian Medical Association reports that between 37-51% of these medical errors are categorized as preventable (2014). Another study lead by Leape and colleagues, further suggests that up to 63% of these mistakes are preventable in the US. In an early survey of physicians (Folkman et al., 1991), 41% of respondents admitted that fatigue was a major factor in their most serious clinical mistake. Cumulatively, these types of medical errors have a financial impact amounting to billions of dollars distributed over hospitals, tax revenues, patients, and the community, thereby putting more stress and strain on the medical system and its workforce (Canadian Medical Association, 2014). Yet, the greatest impact of these errors comes from the resulting loss of life. Indeed, an estimated 440,000 fatalities per year are the result of medical errors in the US (James, 2013).

In addition to increases the risks faced by ill patients, physician fatigue also increases safety risks experience by the physicians themselves. High levels of fatigue have been connected

to twice the likelihood of motor vehicle accidents among physicians as well as a 61% increase in risk for needle related accidents in comparison to low levels fatigue (Eddy, 2005). Further, medical residents working extended hours and on-call shifts are at greater risk for pathogen exposure through needle-prick injuries in comparison to those working a regular shift. It is common for residents to experience greater levels of fatigue during on-call shifts in comparison to regular work hours, and indeed, fatigue better predicts needle-related accidents during on call-shifts in comparison to those sustained during regular work hours (Ayas, Barger, & Cade, 2006; Parks, Yetmen, McNeese, Burau, & Smolensky, 2000). Furthermore, among physicians, fatigue is linked to increased familial stress, increased rates of depression, and reduced feelings of overall wellbeing (Majekodunmi & Landrigan, 2012; Eddy, 2005).

The focus of on-call work in the medical profession exacerbates both the potential for fatigue, as well as its associated risks for both patients and physicians. The sleep deprivation associated with regular on-call shifts has been linked to increases in depression and anxiety, as well as periods of anger and hostility (Krueger & Halperin, 2010; Haines, Marchand, Rousseau, & Demers, 2008; Eastridge, Hamilton, O'Keefe, Rege, & Vaentine, 2003). Shift-work itself has been linked to interrupted sleep patterns, aggravated underlying medical conditions, increased risk of cardiovascular, gastrointestinal and reproductive dysfunction (Knutsson & Boggild, 2010; Nicol & Botterill, 2004) elevated risk of breast cancer (Schernhammer, Laden, Speizer, Willet,

Hunter, & Kawachi, 2003), asthma, diabetes, and epilepsy (Shields, 2002). Higher rates of burnout, emotional distress, emotional exhaustion, job-stress, stomach problems, headaches, and insomnia have also been linked to both shift-work and on-call work, commonly required in the medical profession (Jamal, 2004). Given the relationship between fatigue and shift-work, and between fatigue and on-call work, it is plausible to predict that chronic fatigue, in combination with biological factors, is a contributor to these negative effects for which medical practitioners are at risk.

Physicians are no doubt one of the largest occupational groups facing requirements of long-duration and on call shift work (Pattani, Wi, & Dhalla, 2014), and their resulting fatigue leads to mistakes in them medical environment (Aidman et al., 2015; Fischer, de Castro Moreno, da Silva Borges, & Louzada, 2000). The National Steering Committee on Resident Duty Hours (NSCRDH; 2013) states that 70% of Canadian physicians are subjected to high rates of recurring sleep debt due to work scheduling. The number of medical errors made by physicians has been shown to increase the longer they carry sleep shift-related sleep debt (Gaba & Howard, 2002). As well, physician error rates increase with the number of extended shifts they take on in a workweek (Barger et al., 2006).

1.4.1 Systems and Policies.

Evidence for fatigue as a safety hazard has resulted in the development of policies on

prevention of fatigue related accidents and injuries (Dawson & McCulloch, 2005; Rhodes & Gil, 2002; Standards Australia, 2001; Baker, 2000; Institutes BC, 1999; Mahon & Cross, 1999; Gander, Waite, McKay, Seal, & Millar, 1998; *Occupational Health and Safety Act, in R. S. O. 1990*; *Occupational Health and Safety Management Systems (OHSAS): 18001, Canada*). Both industry and government have contributed to regulations that aim to ensure rest and recovery for employees in various sectors across multiple countries (Rhodes & Gil, 2002; Coplen & Sussman, 2001; Mahon & Cross, 1999; Gander, Waite, McKay, Seal & Miller, 1998; Standards Australia, 2001, Institutes BC, 1999). The most widespread form of these fatigue-related safety regulations are prescriptive hours of service (POS; Dawson & McCulloch, 2005; McCulloch, Fletcher, & Dawson, 2003; Rhodes & Gil, 2002; Burgess-Limerick & Bowen-Rotsaert, 2002; Queensland Transport, 2001). Prescriptive hours of service regulations are rules or laws that control the organization of employees' shifts and schedules. For example, maximum shift length (Lowden, Kecklund, Axelsson, & Akerstedt, 1998; Schroeder, Rosa, & Witt, 1998), minimum break periods (Akerstedt, Kecklund, Lowden, & Axelsson, 2000)**,** and caps for number of sequential shifts (Barton, Spelten, Totterdell, Smith, & Folkard, 1995) are common components to POS systems. In the medical field, patient safety and physician wellbeing are key factors behind the implementation of POS systems (Leape, Brennan, Laaird, Lawthers, Logalio, & Barns, 1991). Indeed, POS regulations are the common form of fatigue management in other high-risk

industries including aviation, transportation, and departments of national defense (Dawson & McCulloch, 2005).

Given that doctors in training are particularly susceptible to both the commission of errors, as well as the development of fatigue (Goldacre, Lambert, & Syirko, 2014; Draper & Louw, 2012; Zukas, & Quinton, Roberts, 2011; Ochsmann, Drexler, Schmid, 2011; Haist, Jacovino, Raymond, & Mee, 2011), special attention has been paid to the development of POS regulations to in an attempt to reduce these risks within medical residents (Canadian Medical Protective Association, 2013). The National Steering Committee on Resident Duty Hours (2013) presently recommends workweeks be limited to 70 hours for all medical residents in Canada. However, under some circumstances, workweeks of up to 100 hours are allowable. Additionally, shift duration is capped at 26 hours, and the allowed number of on-call shifts per week is limited to 4 (Pattani et al., 2014). Similarly, Accreditation Council for Graduate Medical Education (ACGME; 2011) caps allowable schedules at 80 hours per week, with a limit of 30 hours per shift for medical residents practicing in the US. For doctors in their first year of a medical residency in in the US, the maximum shift length is 16 hour (Rosenbaum & Lamas, 2012; ACGME, 2011; Ulmer, Wolman, & Johns, 2008). While medical residents may be more susceptible to fatigue-related errors, experienced physicians are not immune. Presently, the European Time Directive (2009) restricts both medical residents and experienced physicians to a

workweek of only 48 hours, but provides no further recommendations for shift duration and schedules.

Given the above recommendations for work hours among medical professionals, one might expect that the safety risks associated with fatigue would not take effect within a 16-hour, or even a 21-hour shift. However, in a review of the literature on shift duration and safety, Folkard and Tucker (2003) found that the risk of workplace accidents begins to increase non-linearly following the 8th hour of work. For example, they reported that the risk for accident during the 12th hour of a shift is more then double the risk during the 8th hour. Further, Folkard and Tucker reported a greater risk for accidents during nightshifts in comparison to morning or afternoon shifts when controlling for shift length and activities. The authors found that there was a substantial increase in accident risk with each consecutive shift worked overnight, but not for dayshifts. Indeed, the outcome of Folkard and Tucker's (2003) review suggests that the risks to personal and patient safety begin to climb well before the end of a maximum shift for medical residents. This further supports the need for separate regulations night shifts in comparison to day shifts. Similarly, Eastridge and colleagues (2003) investigated the effects of shift work on safety in a medical-specific environment and reported that physicians demonstrated more than double the number of attentional failures during nightshifts as dayshifts. Despite the above evidence, few agencies prescribe differing work regulations for nightshift than dayshift work.

Indeed, in the following section of this proposal I will elaborate on these, and additional, limitations of POS as a strategy for fatigue-risk reduction in the medical field.

1.4.2 System Failure.

The used of POS systems as a tool to mitigate fatigue-related risks is based on the assumption that a set amount of time between shifts allows individuals to recover and return to work alert (NSCRDH 2013, Dawson & McCulloch). Yet, research suggests this is not the case. For example, in a review by Akerstedt and colleagues (2003), the authors reported that 2 consecutive rest days were typically sufficient for employee recovery, but only for those who worked regular daytime schedules of five 8-hour shifts. In comparison, individuals who worked regular night shifts or who worked on an alternative or irregular schedule, required more rest time—up to 4 days—in order to return to work feeling alert. Barton and colleagues (1995) reported a similar outcome when surveying a sample of nurses. In this case, the nurses completed a battery of subjective assessments and cognitive tests in regular intervals over a 28-day period. Overall, a greater decline in cognitive performance was demonstrated over the course of night shifts in comparison to dayshifts and this impairment persisted well into the following rest period. Thus, there is evidence to suggest that the time at which shifts take place is an important factor in the development of workplace fatigue, in addition to shift duration, that is not well accounted for by standard POS systems.

While it is possible for agencies to develop alternative prescriptive hours for night, or alternating shifts, in comparison to regular day schedules, such programs would still be unable to account for differences in health, personality, and circumstance that impact the development of fatigue and its associated cognitive impairments, as well as post-shift recovery (Dawson & McCulloch, 2003). There are many factors that contribute to the state of mental fatigue, above and beyond hours of sleep and rest (Gawron, French, Funke, Hancock, & Desmond, 2001; Van Dongen, Baynard, Maislin, & Dinges, 2004). For example, an 8-hour rest break between shifts does not result in the same sleep quality or mental recovery if it occurs during daylight hours in comparison to night time hours (Dawson & Fletcher, 2001). On-call shifts in particular are associated with reduced attention and vigilance, therefore increasing risk for medical errors, regardless of shift duration or timing, in comparison to regularly scheduled shifts (Dawson & McCulloch; 2005). Yet, neither the POS in Canada or the US provide more conservative regulations for hours worked during on-call shifts versus during regular shifts.

While the relationship between sleep and fatigue is not straightforward, the complexity of the relationship between fatigue and performance further limits the effectiveness of POS systems. Indeed, the impact of fatigue on cognitive and motor performance is mediated by trait-like vulnerabilities within individuals (Van Dongen, Baynard, & Maislin, 2004). To put it simply, not everyone becomes fatigued, or experiences fatigue in the same way and POS systems

cannot take this into account. Additionally, differences in environment (e.g., specialty, workplace) further contribute to the ineffectiveness of this "one size fits all" approach to reducing fatigue risk. Due to variations in geographic location, and medical specialty, specific policies and regulations will not be applicable or effective for all medical staffers. As an example, a family physician working in a rural and remote area may face different stressors in relation to fatigue in comparison to an endocrinologist working in a highly population urban area.

Not surprisingly, there is little evidence supporting an effect of fatigue-reduction based on prescriptive hours of service implementations. Similarly, there is little support for the expected improvement of POS systems on patient safety risks (Philbert, Nasca, Brigham, & Shapiro, 2007). Systematic reviews have been unable conclude that the POS systems relate to improved clinical outcomes (Moonesinghe, Lowery, Shahi, Millen, & Beard, 2011), and report no improvements, or only minimal improvements in patient safety since POS regulations were adopted in the US or in Canada (Drolet, Sangisetty, Tracy, & Cioffi, 2013; Antiel, Reed, Van Arenodonk, Wightman, Hall, & Porterfield, 2013; Vlopp, Rosen, & Rosenbaum, 2007).

In summary, it is clear that POS regulations alone are not sufficient to address fatigue among physicians. There are too many factors contributing to fatigue that go above and beyond sleep history and work hours. As discussed above, shift variability and time, as well as volume of

work can increase fatigue and thus lead to medical errors and these factors are pervasive in the medical profession. Indeed, the American Medical Association has called for shift from POS to individual monitoring and responsibility. Recently, they have released an online tool to monitor fatigue risk in physicians using self-reported fatigue, sleep, workweek, and wellness information. Contrastingly, the National Steering Committee on Resident Duty Hours has stated that self-report measures are unreliable in populations of medical professionals due to the pressures and competitiveness associated with working in the medical field (2013). Thus, the development of new objective assessments may improve fatigue management systems developed for medical professionals.

1.5 Assessing Fatigue

As evidenced in section 1.4.1, the current systems in place to manage fatigue risk in medical professionals have been shown to be ineffective. Indeed, focus has been called to the need for individual-based monitoring of fatigue rather than top-down schedule regulations. Yet, in order to assess fatigue acutely, and on an individualised level, tools for fatigue measurement must be employed. Presently a variety of measures exist within the field that have been validated as fatigue assessments. Yet due the potential unreliability of self-report measures, as well as technological limitations of objective measures, few have been deemed feasible for use in a medical environment in order to mitigate fatigue-related risk. In the following sections I will

provide a brief review of popular fatigue measures currently available, as well as their associated limitations.

1.5.1 Subjective Measures.

A variety of psychological tools are used to assess fatigue for both research and clinical purposes. Presently, all validated measures of fatigue assessment in particular, take the form of self-report questionnaires (for a review see: Shahid, Shen, & Shapiro, 2010). Unfortunately, the majority of work that has been done to develop and validate these measures has occurred in the area of chronic illness and patient care as a means to assess psychological effects of disease progression (Egerton et al., 2015; Dittner, Wessely, & Brown, 2004; Shapiro et al., 2002) thus limiting the validity of these tools when used in healthy populations (Shahid, Shen & Shapiro, 2010). Currently, the most commonly used fatigue assessments include the Fatigue Severity Scale (FSS; Krupp, LaRocca, Muir-Nash, Steinberg, 1989), the Chalder Fatigue Scale (CFS; Chadler, Berelowitz, Pawlikowska, Watts, Wessley, Wright, et al., 1993), and the Fatigue Impact Scale (FIS; Fisk, Ritvo, Ross, Haase, Marrie, Schlech, 1994).

It is important to note both the Stanford Sleepiness Scale (Hoddes, Zarcone, Smythe, Phillips, Dement, 1973) and Epworth Sleepiness Scale (Johns, 1994) are also popular assessments used in fatigue research. However, these scales primarily assess, and are validated for measuring sleepiness. Sleep reduction or deprivation is one factor that can contribute to

fatigue; however, the mental state of fatigue itself does not directly result from lack of sleep, and is not interchangeable with the construct of "sleepiness." (Ramsay, 2000). This is an important distinction, and as such I will refrain from reviewing sleepiness questionnaires, as they would not be applicable to physician fatigue.

Beyond the confusion between the constructs of sleepiness and fatigue, there are additional issues of methodology throughout the field of fatigue research. As cautioned in Shahid and colleagues' (2010) review of both sleepiness and fatigue measures, scales for the assessment of fatigue are designed to collect information about specific *aspects of fatigue*, as well as the impact of fatigue on a patients, or research subject's life. Thus, while individual measures may provide good interrater reliability, and even construct validity, there is often a great deal of variation in the predictions and assumptions that one can make from the results of these self-report questionnaires.

As an example, the Fatigue Severity Scale (FSS) asks respondents questions about how fatigue affects their daily life, in order to assess the severity chronic-like fatigue patterns. For example, "My motivation is lowered when I am fatigued" and "Fatigue causes frequent problems for me." It is plausible that such information may be helpful in monitoring chronic fatigue, and its impacts in medical professions. However, this measure does not provide a quantified representation of *acute* fatigue—that is, it cannot tell us if a medical practitioner is experiencing

low, moderate, or high fatigue at a specific time. Indeed, the FSS has been shown to predict treatment groups in chronically fatigued patients, as well as changes in fatigue patterns over time (Hossain, Reinish, Kayumov, Bhuyia, Shapiro, 2003; Krupp et al., 1989). It has not however, been validated to assess acute fatigue, or to predict errors or performance between fatigue conditions in individuals. Similar to the use of the FSS, the Chalder Fatigue Scale (CFS; Chadler et al., 1993) is also primarily used to assess chronically fatigued individuals. While it is not unlikely that many physicians experience chronic fatigue, the primary need in the field of occupational health and safety at this time is the ability to measure the severity of fatigue in individuals at a specific time point in order to predict the risk associated with their continued performance.

The Fatigue Impact Scale, the last of the commonly employed fatigue scales identified above, is the only one to include questions pertaining to current, acute fatigue (Fisk et al., 1994). Thus, this scale offers the most potential as a tool to predict safety risks in medical workplaces through physician fatigue assessment. Unfortunately, only a limited number of items on this scale represent acute fatigue. Furthermore, given that this scale assesses fatigue through its perceived impact on the respondent's cognitive and physical functions, the outcome of the scale is confounded by the individual's accuracy at identifying the deficits her or she is experiencing.

Beyond the individual limitations of these survey measures, there are widespread

limitations to the use of self-report measures to assess physician fatigue as a means to predict safety outcomes in the medical workplace (Lal & Craig, 2007). For example, the potential for discrepancy between an individuals perception of fatigue impact, and what they are exhibiting behaviourally, is a limitation not only in the use of the FIS survey but can also be applied to any measures relying on self-report of subjective feeling and experience. In line with this concern, past literature has indicated that individuals may not be able to accurately identify their own fatigue, or be able to identify when their performance is reduced due to fatigue (Schmidt et al, 2009; Moller, Kayumov, Bulmash, Nhan, & Shapiro, 2006; Phillip et al., 1997; 2003; Belz, Robinson, & Casali, 2004; Lenne, Trigs, & Redman, 1997; Baranski, 2007; Horne & Baulk, 2004; Lisper, Laurell, & Van Loon, 1986; Nordbakke & Sagber, 2007, Phillip et al., 2003). In the case of medical practitioners, pressures to conform and compete, as well as desires to help those in need, may lead individuals to falsely complete survey measures in order to ensure they are not removed from duty. Indeed, the National Steering Committee for Resident Hours (2013) has cautioned against the use of self-report in fatigue monitoring for this very reason.

In sum, the survey measures currently available cannot be used to adequately assess acute physician fatigue in the workplace. A majority of review literature suggests that survey methods be used in combination with other measures to assess fatigue from a perspective of global functioning, rather than as independent determinates of individualized acute fatigue and

it's associated risks (Lal & Craig, 2007; Shahid, Shen, & Shapiro, 2010). Further, reviews of fatigue measurement and research suggest that future studies should move towards developing objective measures of fatigue as the lack of such tools has created both gap in the this field of study, and, substantial limitations in fatigue-related experiments of the correlational, quasi-experimental, and predictive design (Lal & Craig, 2007).

1.5.2 Available Objective Measures.

Given the concerns discussed in section 1.5.1, and the recommendations of reviews on this topic, a large body of research has attempted to measure and identify fatigue in an objective manner. Such attempts include psychological, hormonal, perceptual, and electroencephalographic studies (see Lal & Craig, 2007 for a review). Most often, attempts at developing new, objective tools for the assessment of fatigue have employed the methods of eye tracking, EEG signal, or ERP components. Unfortunately, at this time the available tools for objective fatigue measurement are limited (Lal & Craig).

Presently, the Optalert Alertness Monitoring System (OAMS; Johns, Chapman, Crowely, & Tucker, 2008a) is a common physiological fatigue tool used primarily in the industrial sector. This particular system based on eyelid movement and aims to identify drowsiness levels based on changes in blinking patterns. Additional information regarding the specifications and requirements of the Optalert system can be found at http://www.optalert.com/. Similar to

sleepiness scales, the Optalert system has been used in research on the topic of driver fatigue (Aidman, Chadunow, Johns & Reece, 2013) despite its focus on an alternative construct—drowsiness, or *sleep propensity*. The aim of such studies has been to use drowsiness ratings of the Optalert to identify and predict risk of vehicle crash. Indeed, the Optalert system could be similarly tested to predict risk of medical error based on the drowsiness of physicians and the portability of the Optalert system makes it a potential candidate for on-site assessment of physicians. In fact, this system is already popular among high-risk industries and has been employed in mining (Caterpillar Global Mining, 2008) transportation (Williamson, Lombardi, Folkard, Stutts, Courtney, & Connor, 2011; Smith, Horswill, Chambers, & Wetton, 2009) and aviation sectors.

Unfortunately, while the Optalert assessment shows promise as both a predictor of drowsiness-based error, and as a field-based tool for use in hospitals, there are several limitations of this system that may prevent its use for this purpose. Primarily, while physician drowsiness is associated with both fatigue-related, and independent safety risks (Eddy, 2005; Eastridge et al., 2003), fatigue-related impairments that develop gradually, and prior to the onset of drowsiness (Grandjean, 1979; 1988) are overlooked by this system. In particular, fatigue-related deficits in decision-making and judgment (Leiberman, Tharion, Shukitt-Hale, Speckman, & Tulley, 2002) cannot be identified or predicted by the Optalert system, and yet are critical factors in the risks of

misdiagnosis, trauma, and fatality posed to patients by physician fatigue (Gaba & Howard, 2002, Trejo et al., 2012; Majekodunmi & Landrigan, 2012).

Furthermore, the feasibility of the Optalert system as a tool used in public, or not-for-profit sectors is limited by both the financial overhead and level of expertise required in operating this system. The Optalert system requires constant monitoring by trained staff in order to alert individuals of increases in risk as they approach drowsiness (Aidman et al., 2015, see also: http://www.optalert.com/). While this system could potentially be used to assess acute drowsiness in physicians, its validity as a risk-prediction tool is based on continuous assessment (Aidman, Chadunow, Johns & Reece, 2013; Williamson et al., 2011) that would not be possible in the case of physicians.

Research into additional tools for the objective assessment of fatigue are ongoing, but are presently in exploratory stages of development (see Lal & Craig, 2007 for a review). A series of studies have sought to use the variability of heart rate as an assessment of acute fatigue (Hartley, Fatigue and Driving, 1995; Hartley, Arnold, Smythe, & Hansen, 1994). However, while the research groups reported identifiable changes in heart rate over time, these characteristics of heart rate did not correlate to performance, or to subjective fatigue.

1.5.3 Event Related Potentials and the P300 as a tool for Fatigue Measurement.

Despite the shortcoming of the above studies, some promise has been shown for the

potential for objective fatigue assessments in the field of EEG research, using event-related-components (Boksem, Meijman, & Lorist, 2005, 2006; Kaseda, Jiang, Kurokawa, Mimori, & Nakamura, 1998; Lal & Craig, 2007; Murata, Uetake, & Takasawa, 2005; Polich & Herbst, 2000). In particular, the P300 component (for a review see Polich & Kok, 1995) has been shown to predict fatigue (Zhao, Liu, & Zheng, 2012; Schmidt et al., 2009; Gruoping & Kan, 2009; Uetake & Murata, 2000; Murata et al., 2005), as well as cognitive performance (Kaseda et al., 1998; Luu, Tucker, & Stripling, 2007; Murata et al., 2005; John Polich, 2007; Portin et al., 2000). The P300 component is identified as a positive deflection of the ERP waveform between 300-400 ms post-stimulus onset, and is reliably elicited by the oddball task (Polich 1999), a common experimental task used in psychophysiological research (Picton, 1992).

Characteristics of the P300 such as latency and amplitude have been repeatedly used as markers for cognitive impairment (Polich & Herbst, 2000; Polich, Romine, Sipe, Aung, & Dalessio, 1992; Aminoff & Goodin, 2001; Casanova-Gonzalez, Cabrera-Gomez, Aquino-Cias, Aneiros-Rivas, & Fernandez-Bermudez, 1999; Elger, et al., 2002; Piras, et al., 2003; Pokryszko-Dragan, et al., 2009; Sundgren, et al., 2015). The magnitude of this component scales with the allocation of attentional resources (Kramer & Strayer 1988; Wickens et al., 1993) and it is theorized that the P300 reflects brain activity associated with context monitoring, working memory, and stimulus recognition (Kutas et al., 1977; Polich, 1983, 1986a, 1990b; Donchin et

al., 1986; Emmerson 1990, Polich & Martin 1992)—all of which are factors associated with the specific safety risks of physician fatigue (Eddy, 2005; Eastridge et al., 2003; Gaba & Howard, 2002).

Of significance to the potential effectiveness of P300 characteristics to predict fatigue-related risks in the medical environment, this component also serves reflection of neural activity associated with judgement and decision-making (Donchin et al., 1986)—the same processes important to the commission of medical error, but left un-assessed by alternative objective measures. Thus, this ERP component is linked to a large variety of cognitive processes that are both necessary for medical practice, and significantly impaired during the fatigue state (Wickens et al., 1983, Kramer & Stayer 1988), positioning it as a plausible method to be further developed as an on-site fatigue assessment tool for medical practitioners. Not surprisingly, several reviews in the area report that EEG measures are the *most* promising tool currently available to objectively measure fatigue (Lal & Craig, 2007, Schmidt et al., 2009; Murata et al., 2005; Horne & Reyner, 1995a; Kardi & Vallet, 1994).

1.5.4 Current Barriers to the use of the P300 to assess fatigue in the medical workplace.

While previous studies have supported EEG and the P300 as a potential tool for fatigue assessment (Lal & Craig, 2006, Murata et al., 2005), various limitations prevented such a method

from being usefull in a field-based environment. The need for high data quality in order to meaningfully interpret ERPs (Picton et al., 2000; Luck, 2014) is a major barrier to the use of field-based ERP research, and thus, would limit the feasibility any on-site ERP-based fatigue protocols for physician assessment. Additional specifications recommended for the effective collection and interpretation of ERP data, such as the number and quality of electrodes (Coles, Gratton, Kramer, & Miller, 1986; Kutas, 1997; e.g., Srinivasan, Tucker, & Murias, 1998) increase the cost of data collection, and thus reduce the feasibility the of a P300-based tool for measuring physician fatigue in the workplace. Indeed, research-quality EEG systems can cost upwards of 75,000 USD. Furthermore, the expertise required to properly set up and monitor the EEG data collection from each individual subject would result in the need hire trained professionals, further inflating the cost of any assessments based on ERPs.

A typically experimental set-up in an ERP laboratory usually requires separate computer systems for stimulus presentation and data recording (Luck, 2014), signal amplifiers (Picton et al., 2000), system battery, and individual electrodes that are hard-wired into the system. Not considering the high cost of this method of data collection, there still little doubt that that the typical experimental set-up is not feasible for fieldwork due to its cumbersomeness and lack of portability (Krigolson et al, 2016). Furthermore, experimental set up for a standard 64 channel EEG system is lengthy and can take up to an hour prior to the beginning of the experimental

paradigm. Evidently, a 1-2 hour assessment of fatigue would be impossible in the medical profession, given the magnitude of work, quick changes to schedules, and limited time commonly experienced by on-shift doctors (Phillip et al., 2015). The nature of work in the medical field would also contribute to loss of data, time, and money, as it would likely be a common occurrence for physicians to need to urgently end an assessment. In such a case, the loss to time, data, and financial input would not be recoverable. Thus, while the P300 component has been identified as a promising tool for the objective assessment of fatigue, the validated methods and requirements of ERP research have proven too extensive for use in an on-site assessment for physicians.

1.6 New Methods

As a potential solution to the limitations commonly associated with using EEG as a measurement tool in a real-world environment, I propose the use of new consumer-grade technology in place of a traditional EEG system. The MUSE EEG headband (InteraXon Inc.) is a low-cost, portable EEG system that has recently been identified as a potential low-cost, portable, and effective tool or the collection of ERP data (Krigolson, et al., 2016). Indeed, the limitations associated with the cost and portability of ERP research formed the rational for the validation of the MUSE carried out by Krigolson and colleagues.

In particular, Krigolson and colleagues were able to successfully extract and quantify

characteristics of the P300 component from data collected with the MUSE EEG system during a traditional oddball paradigm (Polich et al., 1995). Specifically, the contrasting conditions in their oddball task successfully elicited a difference in amplitude, consistent with the timing and shape of the P300 component, and furthermore in line with the characteristics of the component observed from comparison analyses using a standard EEG set up (in which the electrode array was reduced to MUSE electrode locations). It is important to note here however, that in both the P300 component extracted from MUSE recordings, as well as from the parallel standard-system recordings, the polarity of the grand average waveforms for condition, and thus of the component were reversed in comparison to the traditional component.

As discussed by the aforementioned research group, the recent validation of the MUSE EEG headband will no doubt open doors for the portability of ERP research (Krigolson et al., Submitted 2016). Based on my review of the tools currently used in the area of fatigue assessment, I believe the MUSE, in combination with a standard oddball task, provides the most plausible pathway for the development of an on-site, objective assessment of physician fatigue. Indeed, many of the limitations associated with ERP field research do not apply to the MUSE system. Currently, the muse headband retails at *bestbuy.ca* for less than $200 CAD, significantly reducing the cost associated with a start-up of a typical EEG system. Furthermore, experimental set up can be completed in less than 1 minute, and only requires the MUSE headband and the

appropriate software downloaded onto a laptop computer, thus increasing its feasibility for field use with busy physicians in the medical setting. Presently, Krigolson and colleagues are working to port their MUSE ERP software to iOS, allowing researchers to conduct MUSE studies with the use of an Apple iPad, further improving the portability and versatility of this new tool.

1.7 The Proposed Study

Through a series of 3 experiments I aimed to assess both the validity and the feasibility of the MUSE EEG system (InteraXon, Toronto, Ontario)—in combination with a basic oddball task (Polich et al., 1995)—as an objective method for assessing fatigue in a field-based medical environment. Importantly, the experimental paradigm itself was a minimal trial, passive oddball task. In order develop a tool that is usable under the tight time constraints of the medical environment, I kept the experimental paradigm to a limited 3-minute duration. The focus of these experiments began with confirmation of the relationship between the P300 and perceived fatigue under standard conditions, and ended with a test of the MUSE assessment in a medical environment in which fatigue was experimentally induced.

CHAPTER TWO: EXPERIMENT 1—CONFIRMATION OF THE P300 AS AN INDICATIOR OF FATIGUE USING STANDARD EEG SYSTEM

2.1 Introduction and Proposal

With this study I aimed to provide evidence in support of the use of a new portable EEG tool, the MUSE, as a measurement for fatigue that can be used in the field, and in particular, in the medical workplace. While reviews on the topic of objective fatigue measurement suggest that EEG, and in particular the P300 component, is a promising avenue for the development of such an assessment (Lal & Craig, 2007; Schmidt et al., 2009; Murata et al., 2005; Zhao et al., 2012) the majority of experiments discussed are dated, and comprised of low sample sizes (e.g., N = 5, Murata & Utake, 2005; N= 7, Kaseda, Jaing & Kuruokawa, 1998; N = 29 Schmidt et al., 2009). Thus, prior to investigating the use of the MUSE system itself to assess fatigue via P300 magnitude, I completed the present experiment to confirm the relationship between subjects' perceived fatigue and P300 by employing a larger sample size in comparison to prior studies with data collected from a standard, research-level, 64 channel EEG system (ActiCHamp system from Brain Vision). This investigation comprises Experiment One.

Additionally, Experiment One was used to confirm that the passive oddball task developed for this study was effective in eliciting the P300 component, prior to using this task in following experiments. I predicted that that the experimental task would be successful in eliciting

the P300, and that the magnitude of this component would be reduced in subjects who perceived themselves as being highly fatigued in comparison to those who did not. Further, I predicted that the magnitude of the P300 component would be negatively correlated with subjects' perceived fatigue.

2.2 Method

2.2.1 Participants.

Thirty-five participants took part in Experiment One. Participants were recruited using the University of Victoria's online experimental sign up system, and completed the study in exchange for bonus points that could be used to increase their grades in select psychology courses offered through the university. Any students with normal, or corrected-to-normal vision and who were able to wear a standard EEG cap were eligible to participate in Experiment One.

Informed written consent was obtained from all participants upon their arrival to the laboratory on the day of the experiment. All experimental procedures were approved by the Human Research Ethics Board at the University of Victoria prior to the beginning of data collection and followed ethical standards prescribed in the 1964 Declaration of Helsinki.

2.2.2 Procedure.

Participants arrived at the Neuroeconomics laboratory and following informed consent reported their perceived fatigue via paper assessment (see section 2.2.4 for assessment details

and rationale) and were then fitted with an electrode cap and 64 electrodes. They then completed the experimental task, a visual oddball task similar to those previously shown to elicit the P300 component (Polich et al., 2000, Polich 1995), and identical to the passive oddball task used by Krigolson and colleagues (2017), while EEG data were recorded. Participants viewed the experimental task on a 24" LCD computer monitor. The experiment took place in a dimly lit, sound dampened room in the Neuroeconomics Laboratory at University of Victoria.

2.2.3 Experimental Task.

The experimental task used to elicit the P300 component was a passive, visual oddball task coded in MATLAB (Version 8.6, Mathworks, Natick, USA) using the Psychophysics Toolbox extension (Brainard, 1997). During each trial of the experiment, subjects were presented with a black fixation cross on screen for 300-500 ms, followed by a green or blue circle, on screen for 800-1200 ms. The variation of stimulus presentation time was incorporated into the task in attempt to maintain attention and reduce habituation in participants (Krigolson et al., 2017). The frequency of presentation for the blue and green circles differed such that the blue circles appeared less frequently (the "oddball": 25%) than the green circles ("distractor": 75%). Participants were instructed to mentally keep track of the number of blue circles (oddballs) within each block of the experiment. Instructions were provided both verbally by the experimenter, and textually (on screen).

The task included 3 blocks of 40 trials. While this is a small number of total trials in comparison to previous studies evaluating the P300 (Polich, 1995; Polich & Herbst, 2007) the goal of the current study is to inform the development of fatigue measures to be used in real-world settings such as medical workplaces, and as such I aimed to elicit and measure the P300 component with a minimum number of trials, in a task no more than 3 minutes in duration.

2.2.4 Perceived Fatigue

Given that I was unable to identify a validated measure of acute fatigue for healthy populations in my literature review (See section 1.5.1), I opted to assess perceived fatigue with a novel and non-validated item (see Appendix A for fatigue measure). Specifically, participants responded to the item "I am mentally tired" on a scale from 1 to 5, 1 being strongly disagree, and 5 being strongly agree. The particular wording of this question was chosen in an attempt to isolate the construct of *acute* fatigue, and to keep the question simple and accessible given the various uses of the term "fatigue" in both research and casual discussion (e.g., physical fatigue, sleepiness, and drowsiness; Ramsay, 2000; Shahid, Shen, & Shapiro, 2010). While the use of a non-validated measure was not ideal, this purpose of Experiment One was merely to inform my future experiments leading up to Experiment Three in which fatigue was experimentally induced. Indeed, the primary rational for the present study in its entirety is that there are no available methods for assessing physician fatigue due to the limitations of existing fatigue measures.

2.2.5 Data Acquisition.

Electroencephalographic data were recorded from 64 electrodes (standard 10-20 layout; ActiCAP, Brainproducts, GmbH, Munich, Germany) using Brain Vision Recorder Software (Version 1.21, Brainproducts, GmbH, Munich, Germany). Electrodes were initially referenced to a common ground, and electrode impedances were be kept below 20 kΩ. Data were sampled at 500 Hz, amplified (ActiCHamp, Revision 2, Brainproducts, GmBH, Munich, Germany), and filtered through an antialiasing low-pass filter of 8 kHz. A DATAPixx stimulus unit (VPixx, Vision Science Solutions, Quebec, Canada) was used to ensure temporal coincidence of event-markers with experimental stimuli.

2.2.6 Data Processing.

Data were be processed offline with Brain Vision Analyzer 2 software (Version 2.1.1, Brain Products, GmbH, Munich, Germany) using standard methods from the Neuroeconomics laboratory (available at: http://www.neuroeconlab.com/data-analysis.html). Excessively noisy electrodes were removed and EEG data was then re-referenced to an average of the mastoid electrodes. Data were filtered using a dual pass Butterworth filter with passband of 0.1 Hz to 30 Hz as well as a 60 Hz notch filter. This step was used in order to remove artifacts in the data caused by surround in electrical interference and muscle movement (Luck, 2014). Data epochs of 3000 ms surrounding each event of interest were extracted from the continuous EEG.

Independent component analysis was employed in order to correct ocular artifacts (Delorme & Makeig, 2004; Luck, 2014). Data were then reconstructed and any previously removed channels were interpolated using spherical splines. New 600 ms epochs were constructed containing 200 ms prior to and following the events of interest (i.e., presentation of blue and green circles). Finally, segments were processed by an artifact rejection algorithm that removed segments with gradients of greater than 10 μV/ms or with a 100 μV absolute difference within the segment. For each participant and event of interest (oddball; control), ERP waveforms were computed by averaging the segmented EEG data for each electrode. Next, I created a difference wave by subtracting the average waveform for the oddball (blue circle) from the distractor (green circle) for each participant. The P300 component was quantified for each participant as the mean of the of the individual difference wave within a 50 ms range calculated around the peak of the grand average difference wave in the time range of the P300 component (300-400ms; Polich et al., 2002).

Averaging all corresponding ERPs across all participants within each group (high or low fatigue, see section 2.2.7) I created grand average waveforms for the 2 conditional waveforms, as well as for the difference waveforms. These grand-average waveforms represented the averaged response to stimuli for each group.

2.2.7 Data Analysis.

Given that participants rate their fatigue on a single item scale, no calculation was required to assess fatigue. Rather, the recorded value for how "mentally tired" participants felt was used in statistical analysis evaluating the relationship between subjective fatigue and P300 magnitude. Data were grouped into high-fatigue and low-fatigue groups based on a median split of fatigue scores. I then confirmed the presence of the P300 component within each group by visual inspection and by conducting a t-test of the peak amplitudes of the difference waves with zero ($\alpha = 0.05$). This is a standard test used to confirm the presence of ERP components used by the Neuroeonomics Laboratory, the logic of this test is that if the component is not present, the amplitudes should be normally distributed around zero, and the test would fail, confirming that there is in fact a difference within the component. In this case the test was used to confirm that the 3-minute oddball task was successful in eliciting the P300 component.

Next I employed an independent samples t-test to confirm a difference in component peaks between high and low fatigue groups. Finally, I assessed the relationship between difference wave peaks and reported fatigue by computing a Pearson R value over all participants.

2.3 Results

In line with previous usage of passive oddball tasks (Krigolson et al., 2017) upon visual analysis, I found my experimental paradigm reliably elicited the P300 component. The peaks of

the average difference waves were different from zero when tested with an independent samples T-test, $t(32) = 12.03$, $p < 0.0000$. When data were binned based on subject's perceived fatigue, the high fatigue group (mean fatigue = 7.19, 95% CI [6.63, 7.75]) demonstrated smaller P300 amplitudes, based on visual analysis, in comparison to their low fatigue counterparts (mean fatigue score = 2.45, 95% CI [2.03, 2.87]). A side-by-side comparison of the waveform for high and low fatigue is presented in Figure 1.

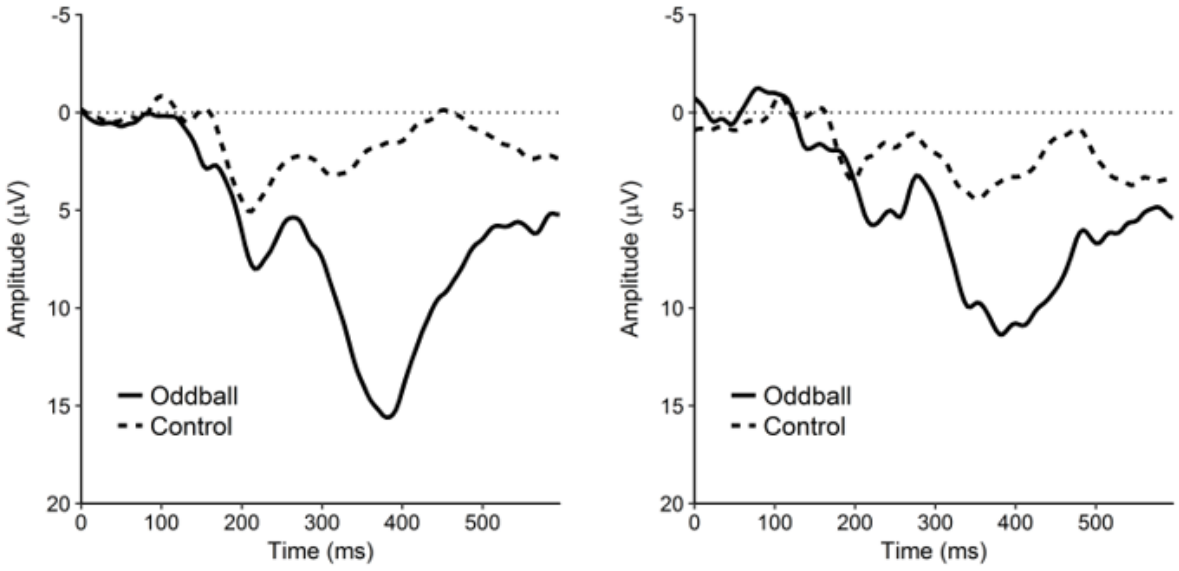


FIGURE 1. Side-by-side comparison of the grand average conditional waveforms for oddball and distractor stimuli for the high and low fatigued groups in Experiment One.

An independent samples t-test of the P300 scores between each group confirmed that the mean peak amplitude of the component was reduced for the high fatigue group (mean peak amplitude = 6.57 uV, df = 15, 95% CI [4.85, 7.95]), in comparison to the low fatigue group (mean peak amplitude = 11.74 uV, df = 17, 95% CI [10.61, 12.86]), t(32) = 4.04, p = 0.0003, Cohen's D = 1.39. A comparison of the component difference waves for high and low fatigue is presented in Figure 2.

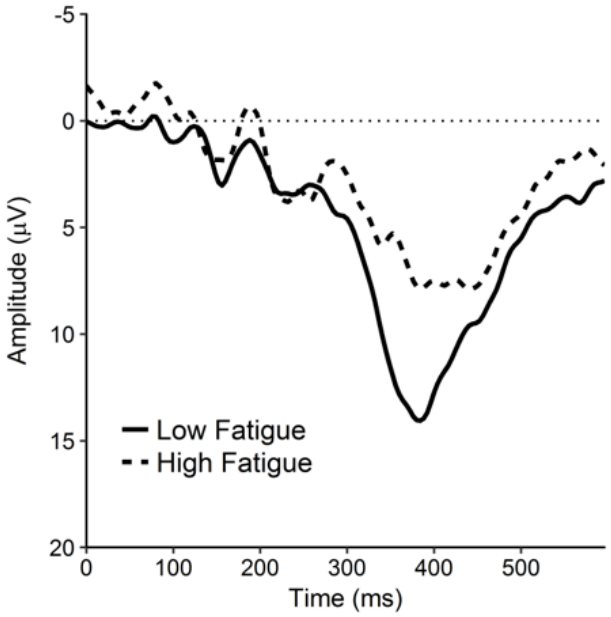


FIGURE 2. Grand average difference waveforms for high and low fatigue groups in Experiment One.

In the overall sample, I found a strong correlation between the severity of perceived fatigue and the peak of the P300 component difference wave, Pearson's r = -0.67, t(32) = -5.05, p = 0.00005. A plot of this relationship is presented in Figure 3.

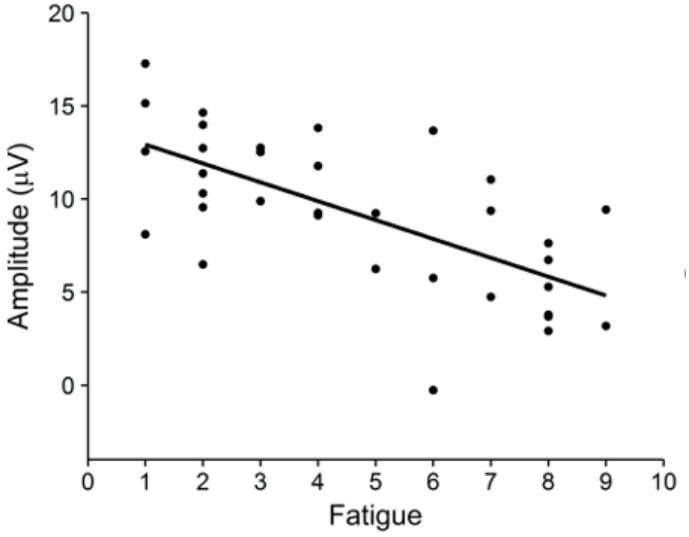


FIGURE 3. Correlation between P300 amplitudes and their perceived fatigue scores for Experiment One.

2.4 Summary

Based on previous literature, some of the major factors reducing the feasibility of ERPs as a field-based fatigue measurement were rooted in portability, cost, and expertise (Krigolson et al. 2017) Indeed, the purpose of this study was to support the development of a *portable* ERP assessment tool for fatigue. While the experiment 1 did not utilize the proposed portable EEG system, it did confirm hypotheses that were necessary for further experimentation. The oddball task I aimed to use for the portable fatigue assessment consisted substantially fewer trials than versions typically used in ERP research (e.g., Boksem et al., 2005; Holroyd & Krigolson, 2007; Krigolson, Hassall, Satel, & Klein, 2015; J Polich & Kok, 1995; Schubert et al., 1998; Williams, Saffer, McCulloch, & Krigolson, 2016), thus with Experiment One sought to confirm that my proposed three-minute paradigm could be used to elicit the P300 component and assess its magnitude under typical experimental conditions—the use of a standard 64 channel EEG system, in a laboratory setting. The results of Experiment One supported my hypothesis in this case and an ERP component with characteristics in line with the P300 was obtained on oddball trials for both high and low fatigued subjects.

Prior to testing a portable EEG system as a potential fatigue assessment, it was important to also test the sensitivity of this component to my chosen fatigue measure in a standard research environment. Again the outcome of experiment one supports my hypothesis. Upon visual

inspection, the P300 component appeared to be blunted, or reduced, for the high fatigue group. Statistical analysis of the difference wave peaks for each group confirmed a difference in mean peak amplitude between groups supporting a reduction in the amplitude of the P300 in the highly fatigued group in comparison to those who reported feeling low fatigue. Further, the results of experiment one confirmed a negative correlation between subjects perceived fatigue ratings and the magnitude of the P300 component, as represented by the peak amplitude of the difference wave, suggesting that as fatigue increases, the magnitude of the P300 is reduced.

The ability to measure P300 magnitude in a short-duration task will greatly improve the feasibility of an EEG tool in assessing fatigue in the field. Medical professionals have limited time and lengthy assessments would put stress not only on individuals but also the systems for schedule, managing, and financing the operations of hospitals and other medical workplaces (National Steering Committee on Resident Duty Hours 2013; Tucker, Bejerot, Kecklund, Aronsson, & Åkerstedt, 2015; Wong & Imrie, 2013). Furthermore, while previous literature supports the use of the P300 as a fatigue measurement, few specific studies have used the P300 as a predictor of fatigue in a controlled experimental setting and the limited body of research in this area is compromised primarily of experiments with low, unfavourable, sample sizes (e.g., Kaseda et al., 1998; Murata et al., 2005; Zhao et al., 2012). As such, Experiment 1 is necessary to replicate and confirm the relationship between the P300 and fatigue, prior to developing new

methodologies based on this theory.

In conclusion, Experiment 1 has provided the groundwork for Experiments Two and Three. It has confirmed the P300 component can be successfully elicited and measured from a passive oddball task, using minimal trials and has provided support for the use of the P300 component as a potential measurement of fatigue.

CHAPTER THREE: EXPERIMENT 2—INVESTIGATION OF P300 AS INDICATOR OF FATIGUE USING THE MUSE PORTABLE EEG SYSTEM

3.1 Introduction and Proposal

Following the confirmation of the relationship between subjective fatigue and P300 magnitude obtained in Experiment 1 I directed my attention to the validation of the experimental paradigm as a fatigue sensitive tool when used on combination with the MUSE EEG system. In order to support my claim of the MUSE as a portable assessment, experiment took place outside of the laboratory, but otherwise followed experimental design outlined in experiment 1.

3.2 Method

3.2.1 Participants.

Seventy-eight subjects took part in Experiment 2. Participants were recruited by word of mouth through the Neuroeconomics Laboratory and through class announcements in the School of Exercise Science, Physical and Health Education. Participating in the experiment took approximately 20 minutes of the subjects' time and they were not compensated with course credits. Again, normal or corrected-to-normal vision was requirement for participation in Experiment Two. Participants completed informed written consent prior to the beginning of the experiment. All experimental procedures were approved by the Human Research Ethics Board at the University of Victoria prior to the beginning of data collection and followed ethical standards

prescribed in the 1964 Declaration of Helsinki.

3.2.3 Procedure.

In Experiment 2, participants completed the same survey measures and experimental task as in Experiment 1. Again, fatigue was assessed based on a single item self report scale asking how "mentally tired" the participants felt. Contrary to Experiment 1, participants completed the experimental task on a 13" Macbook Air Laptop (Apple Inc., California, USA), while EEG data was collected via the MUSE headband. The size of the visual stimuli was adjusted as to be the same as those presented with the standard system, regardless of the reduction in monitor size. Participants completed the task in a quiet location on campus at the University of Victoria.

3.2.4 Data Acquisition.

For Experiment 2 I followed all methodological recommendations for research with the MUSE (InteraXon, Ontario, Canada) put forth by the Neuroeconomics Laboratory. These recommendations were based on the laboratory's previous research with the device, and their development of software tools to be used with the MUSE. An in depth explanation of MUSE methods is available at http://www.neuroeconlab.com/muse.html.

As a summary, EEG data was be acquired via recordings from the 5 electrodes of the MUSE EEG headband. These electrodes are analogous to electrodes AF7, AF8, Fpz, TP9, and TP10 on the standard 10-20 layout. Recording software for the MUSE ran directly from the

Macbook Air Laptop on which participants viewed the experimental task. The MUSE headband was be connected to the laptop via Bluetooth and with the use of muse-io, a library of commands developed by the MUSE team to allow Bluetooth communication through Terminal on Mac OS (details available at: http://developer.choosemuse.com/research-tools/museio). While the MUSE headband comes with onboard signal processing, this feature was be disabled, as it was not compatible with my desired sampling rate of 500Hz.

Quality of incoming signal was assessed on site via custom code written in MATLAB (MATHWORKS, Munich, Germany) by the Neuroeconomics Laboratory and available at http://www.neuroeconlab.com/muse-data-collection.html. This colour codes signal quality by calculating the variance of the data over 500 samples (1 second) and displaying it onscreen for each channel. Additionally, the data stream was visualised on screen as an individual waveform for each channel, which turned colour from red, to orange, to green as variance decreases. Green waveforms represented desirable signal quality with a variance of less than 200. This threshold is somewhat arbitrary, but when piloted, resulted in desirable data recordings, and furthermore ensures blinks will be coded red, for "bad data". Prior to the start of the task, and recording of EEG data, the quality of the data was "accepted" by the research assistant using a key command.

3.2.5 Data Processing.

Again, all methods for analysis of the MUSE data in experiment 2 followed the

recommendations developed by the Neuroeconomics Lab and published in Krigolson et al., 2017. Analyses were completed in MATLAB [Version 8.6, Mathworks, Natick, USA] using specific MATLAB and EEGLAB (Delorme & Makeig, 2004) code, as well as the Neuroeconomics Lab's custom software as discussed above.

The EEG data was first filtered with a dual pass Butterworth filter with a pass band of 0.1 Hz to 15 Hz as well as a 60 Hz notch filter. To assess the ERPs within the EEG data, I used only the pooled data of TP9 and TP10, and removed Fpz, from the data set. Given that the EEG data recorded is referenced to Fpz, the activity at this electrode forms the baseline against which other electrodes are subtracted and thus the amplitude at this site should be zero. Krigolson and colleagues (2016) verified that this is the best method for characterizing the P300 component using this experimental set up.

Data were baseline corrected using the 200 ms prior to stimulus onset. I then extracted segments consisting of the 600 ms following stimulus presentation (the green and blue circles) from the pooled continuous EEG at TP9 and TP10. Here, any epochs with a gradients greater than 10 μV/ms or with a 100 μV absolute difference within the segment were discarded. Given the limited array of the MUSE headband, and my decision to use only electrodes TP9 and TP10 in my analyses, I was not able to remove specific channels due to excess noise. Instead, in cases of extreme noisiness at either of these channel locations (i.e., if > 80% of segments are

discarded) the participant's data was removed from the analyses. This resulted in a loss of data from 3 participants in the sample. Next.

In line with my methods for Experiment 1, average waveforms were then created for each subject and condition (oddball verses control) and individual difference waves were computed by subtracting the average waveform for the oddball condition, from the average waveform for the control condition, for each participant. Grand average waveforms were then computed for each condition and group (i.e., oddball verses control, high verses low fatigue) by averaging across all subjects in each group. Similarly, the difference waves for all subjects in each group were averaged to create a grand average difference wave for each group.

I then used the peak of this difference wave to determine the time interval in which to compute average peaks for each participant as a means of quantification of their associated P300 magnitude (355 ms – 405 ms). Given that there were expected, unavoidable timing lags in data recording due to the MUSE's Bluetooth connection, grand average difference wave peaks for the P300 appeared later in the data segments in comparison to those from standard system. The P300 magnitude was quantified for each individual as the mean peak within a 50 ms interval around the peak of the grand average difference wave that was present in this time period.

3.2.6 Data Analysis

I first confirmed the presence of the P300 component by visual analysis and by

conducting a two sample t-test test of the peak amplitudes of the difference waves with zero ($\alpha = 0.05$). Again, if no component was present, the test against zero will would fail. I then performed a median split of the P300 data based on subjects' perceived fatigue score, creating two groups—high and low fatigue. I then computed grand average ERP waveforms for oddball and distractors for each group, and difference waves for each group. To confirm a difference in the magnitude of the P300 between groups I performed a Two-Sample t-test on the peak amplitudes of the difference waves for subjects in each group. Further, I computed a Pearson R correlation score between fatigue score and peak amplitudes across the whole sample.

3.3 Results

In line with the outcome of Experiment One my experimental paradigm elicited an ERP component consistent with the P300 when used with the MUSE-EEG system. A comparison of difference wave peaks to zero confirmed the presence of the component, $t(74) = 10.52$, $p < 0.0000$.

A median split of the data based on fatigue score resulted in a high fatigue group with a mean score of 6.45, 95% CI [6.11, 6.77], and a size of $n = 38$, and a low fatigue group with a mean score 2.92, 95% CI [2.64, 3.20], and a size of $n = 37$. In line with my predictions, and with the outcome of Experiment One, and with my principal hypothesis, I found that the peaks of the component difference waves were reduced for the high fatigue group (mean peak amplitude =

1.50 uV, 95% CI [1.07, 1.94]) in comparison to the low fatigue group (mean peak amplitude = 3.22 uV, 95% CI [2.69, 3.75]; t(73) = 4.27, p = 0.0005, Cohen's d = 0.98). A side-by-side comparison of the ERP waveform for each group is presented in figure 4, and a comparison of the difference waves of the components for each group is presented in Figure 5.

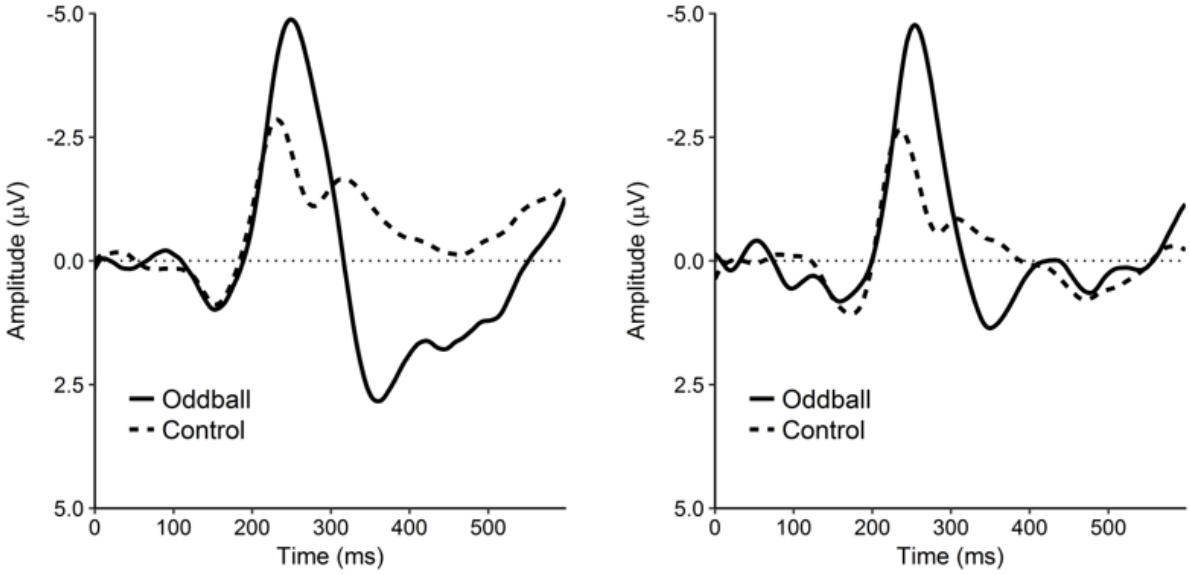


FIGURE 4. Comparison of conditional waveforms for oddball and distractor stimuli for the high and low fatigued groups in Experiment Two.

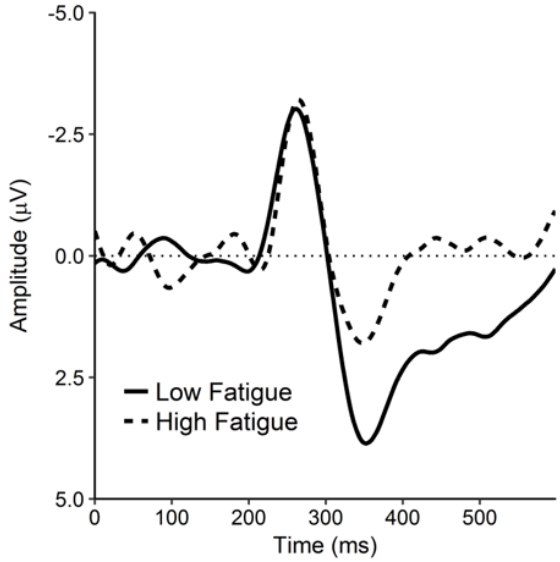


FIGURE 5. P300 difference waves for high and low fatigue groups in Experiment Two. Calculated as the average of all oddball-distractor difference waves for subjects in each group.

Similar to the outcome of experiment 1, I found a strong correlation between the severity of perceived fatigue and the peak of the component difference wave (Pearson's r = -0.62, t(73) = -6.49, $p < 0.00000$. This relationship is plotted Figure 6.

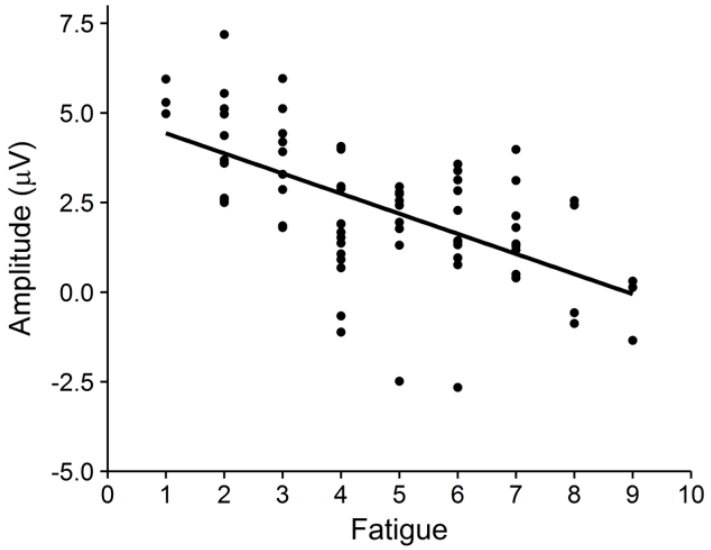


FIGURE 6. Correlation between P300 amplitude and perceived fatigue scores for Experiment Two.

3.4 Summary

Experiment Two was the first experiment to use the MUSE EEG headband in combination with the 3-minute P300 eliciting task. Its purpose was to further confirm the relationship between the P300 and perceived fatigue when data were collected with a MUSE headband in an non-laboratory environment rather than a standard EEG system and experimental set up. Thus, a secondary purpose of Experiment Two was to allow qualitative assessment of the feasibility of conducting a fatigue assessment with the MUSE in a non-experimental environment, as this would be necessary for any real world application of the MUSE fatigue assessment in the medical environment, as well as for the procedures of Experiment Three.

Again, my hypotheses for Experiment Two were confirmed by my analysis of the data. The presence of the P300 component in both high and low fatigued groups, as confirmed by visual inspection, and statistical testing, supports my prediction that the MUSE can be used in combination with the short oddball task to elicit and quantify the P300 component when used in a field-based setting. Further, the difference in peak amplitudes of the difference wave between groups suggests that the magnitude of the P300 component is reduced in subjects who are highly fatigued in comparison to those who are not, when categorized based on a median split of a subjective, single-item measure of their perceived fatigue. Further, the negative correlation between peak amplitudes of the difference waves and fatigue scores further supports the outcome of experiment 1, suggesting that the magnitude of the P300 component is further reduced as fatigue severity increases. Given that the results of Experiment Two are in line with those obtained from experiment 1, when using a laboratory-based procedure, I concluded that Experiment Two supports the use of the MUSE system and the short oddball task as an ERP assessment that can be used in a field environment that is sensitive to individual's perceived fatigue.

CHAPTER 4: EXPERIMENT 3—USING THE PORTABLE MUSE SET-UP TO INVESTIGATE FATIGUE IN A SIMULATED MEDICAL ENVIRONMENT

4.1 Introduction and Proposal

While confirmation of the MUSE assessment's sensitivity to perceive fatigue was a first step towards validating this tool as a fatigue measure, it was imperative to test the assessment in a medical environment using a sample similar to the medical professions that would utilize the tool in a real work setting. Indeed, while experiment 2 supported effectiveness of the combined use of the MUSE EEG headband with short oddball task as an ERP measure sensitive to fatigue, the sample consisted of undergraduate university students whose personality traits and experience with fatigue as a population may differ from a sample of physicians. Thus, in experiment 3, I utilized a sample of senior year medical students and junior medical residents. While in experiment 2, the MUSE assessment was completed outside of the laboratory, the environment of the assessment may still have been more controlled than a real-world hospital environment where participants are under high levels of stress. Thus, experiment 3 took place in a simulated medical environment. The foundation of this environment were existing medical training laboratories located at teaching hospitals affiliated with University of Victoria, and University of British Columbia, and University of Calgary medical schools. Importantly, the purpose of these training facilities is to mimic the real life experiences of a physician working

on-call in a hospital (Okuda et al., 2009). Such simulations are widely used in the practice of medical education for both students and residents, and have been shown to effectively represent real-world settings in order to prepare students for the medical field (Okuda, et al., 2009; Rogers, McConnell, Jones de Rooy, Ellem, & Lombard 2014; Lumley, 2013; Schwind, Boehler, Markwell, Williams, & Brenner, 2011).

In addition to validating the MUSE for fatigue sensitivity in a medical sample and environment, experiment 3 also served the purpose of providing the first evidence supporting a MUSE fatigue assessment using an experimental design rather than the quasi-experimental, correlational designs utilized in experiments 1 and 2. Indeed, two key limitations to experiments 1 and 2 were a lack of experimental manipulation of fatigue, and further, the use of a non-validated subjective measure for rating fatigue. Given both the lack of appropriate measures for identifying acute fatigue in healthy subjects (Sahid & Shapiro, 2010), in addition to the evidence that commonly used subjective measures are insufficient for identifying fatigue for safety purposes, experimental manipulation of fatigue was of utmost importance in assessing the feasibility of the MUSE and accompanying oddball task as a potential fatigue assessment. Indeed, the mounting evidence that individuals poorly evaluate the effects of fatigue on their own cognitive processes (Belz, Robinson, & Casali, 2004; Lenne, Triggs, & Redman, 1997; Moller, Kayumov, Bulmash, Nhan, & Shapiro, 2006; Phillip, et al., 1997; 2003) underlies the

purpose for this study and its aim to develop a new, objective measure for fatigue assessment.

In experiment 3, the use of the medical training environments was used mimic the stress and fatigue associated with a physician's shift as closely as possible. The simulations were comprised of 12-hour shifts in which subjects completed medical cases with the aid of computerized training mannequins, and confederates playing the roles of nurses and attending physicians. Previous literature supported the assumption that participants would experience mental fatigue due to their participation in the simulation (Folkard & Tucker 2003). I predicted that subjects would report greater perceived fatigue at the end of the simulation in comparison to the start, and further, that the amplitude of the P300 ERP component would be reduced post simulation in comparison to pre-simulation.

Here, I sought to validate the use of portable EEG to assess fatigue in a simulated medical environment. The outcome of this experiment further confirmed that the P300 component is sensitive to fatigue, and in particular, that the P300 component was reduced within individuals following a fatigue-inducing clinical simulation. Thus, Experiment Three provides the necessary evidence to support the use of the MUSE and oddball task as a fatigue-sensitive measure to be used in a hospital setting, and to set the ground work for recommended follow-up studies to pilot the technology in a real world setting with on-shift physicians.

4.2 Method

4.2.1 Participants.

Participants were final year medical students, and first year medical residents from the University of Victoria, the University of British Columbia, and the University of Calgary medical school programs. Participants were be recruited via an email sent to all medical students and residents in the area of internal medicine in each of these three institutions, on behalf of the medical sciences department of each respective institution. Experiment Three consisted of an in-depth medical training simulation in which participants took part, under the direction of senior residents and licensed physicians. Participants were not compensated monetarily but volunteered to partake in the study to gain valuable hands-on experience in the medical field. Due to the large time and work commitment of taking part in this experiment, (a minimum of 12 hours) I expected small sample size to be a key limitation in this portion of the study. I proposed, and aimed to enrol 20 subjects in Experiment Three, however only 13 individuals participated. Enrolment in the study operated on a first come first serve basis. Participants provided informed written consent prior to the start of the experiment. All methods for Experiment Three were approved the ethical bodies associated with University of British Columbia, University of Calgary, and Island Health.

4.2.2 Procedure.

While data for experiment three was collected across three different institutions, the procedure for the experiment was the same in all locations. All methods discussed in this section were be replicated at all three institutions, and all three institutions were equipped with the same equipment necessary to carry out the experiment in the same way.

Participants were invited to take part in a 12-hour medical training exercise designed to simulate a day, or night, as an on-call medical resident. The experiment took place in existing simulation environments operated by the medical schools at each of three medical campuses (University of Victoria, University of British Columbia, University of Calgary). Simulation labs at each of the institutions were equipped with high-fidelity medical simulation mannequins as well as medical equipment that would commonly be present in a hospital's emergency room. These simulation labs were used to create realistic case studies in which participants will practice medical skill as well as decision-making. A description of the simulation labs located within Island Health at the University of Victoria can be found at: http://www.viha.ca/professionals/simulation/labs/default.htm.

At both the beginning and end of the 12-hour training simulation, participants completed the 3-minute oddball task used in all previous experiments as well as the single item measure of their perceived fatigue. In addition, participants completed a modified version of the

Occupational Fatigue Exhaustion Recovery scale (M-OFER; Winwood et al., 2005) at these time points. This questionnaire is designed to assess the life impact of *overall* fatigue in working professionals but does contain a subscale for the impact of acute fatigue (items 6-10, Rahmen et al., 2017). Thus the inclusion of this questionnaire was to gain information about participant's perceptions of the impact of their fatigue following the simulation. An example of the M-OFER is presented in Appendix B. EEG data were recorded using the MUSE EEG system, using the exact protocol used in Experiment Two. Additionally, participants completed the single item fatigue measurement used in both Experiments One and Two. The only deviation from the previous methodology is that participants in Experiment Three is that subjects completed the EEG portion of the in the medical simulation environment rather than in a quiet location at the University of Victoria campus.

During the experiment the simulation was monitored and run by research facilitators. These facilitators were hospital and medical-school staff that were trained in how to operate the simulation labs for training exercises. The facilitators were responsible for carrying out the case studies with the participants in Experiment Three. Each participant was assigned to an available-time slot in which the experiment was scheduled. At this time they met with their respective facilitator at the simulation lab for an overview of the equipment and to complete informed consent.

Once the experiment began, the participant was provided with a pager and shown to an on-call room within the hospital. The participant was alerted by pager to a patient situation (case) at four different times during the simulation. Cases involved the need for participants to wake themselves if asleep and move to the case room to evaluate and treat a simulated patient (the SimMan3G; Laerdal Toronto, Ontario). The times at which the cases occurred were randomly assigned in order to mimic a real life experience as an on call physician as closely as possible.

When the participant was paged to see a patient they moved to the simulation lab and were provide with a patient file. The file contained patient medical records in hospital format as well as an overview of the patient's current symptoms and complains. The facilitator also verbally summarized the information contained in the patient file. When the participant finished assessing and treating his or her patient, he or she will was allowed to return to their call room with their pager. While participants were in their call room they were free to spend their time as they chose (e.g., they may read or study) but we will suggested to them that they use this time for sleep. Immediately following the completion of the experiment, participants were given the opportunity to discuss their performance in the simulation with the facilitator for up to 60-minutes. This opportunity will was offered to participants as a means of debriefing, and mitigating any feelings of unease over performance, as well as to improve their learning experience gained from study participation.

4.2.3 Data Processing.

As discussed in previous sections, EEG data were recorded from participants using the MUSE EEG headband and associated 3-minute oddball task. However, in Experiment Three I collected this data at two time points per participant, pre- and post-simulation. All data acquisition and processing methods for the EEG data in experiment 3 were identical to those employed in experiment 2 and described in above sections.

4.2.4 Data Analyses.

In line with previous analysis, I first sought to confirm the presence of the P300 component by both visual inspection and by use of a two-sample t-test of the amplitudes of peaks of the difference waves against zero. Next, I employed a paired t-test to confirm differences in the amplitudes of the difference wave peaks before and after the simulation. Finally I conducted a Pearson correlation of subjects' perceived fatigue and the amplitudes of the peaks of the difference waves for the ERP component.

4.3 Results

A paired t-test of subjects' fatigue scores pre- and post-simulation confirmed an increase in perceived fatigue at the end of the simulation ($t(12) = -3.34$, $p = 0.0059$, Cohen's $d = -1.36$; mean perceived fatigue pre- = 3.69, 95% CI [3.11, 4.28]; mean perceived fatigue post- = 5.69, 95% CI = 4.85, 6.53]). No difference was found in subjects' scores from the M-OFER (two

sample t-test, t(24) = -1.37, p = 0.18; pre-mean = 4.54, 95% CI [3.42, 5.65]; post-mean = 5.85, 95% CI [4.55, 7.14].

The mean P300 amplitude was 4.63 uV prior to the simulation (95% CI [2.79, 6.47]) and 1.34 uV post simulation (95% CI [0.09, 2.59]). A two-sample t-test of component difference wave peaks against zero confirmed the presence of the P300 component in the ERP data collected prior to the medical simulation, t(12) = 4.50, p = 0.0007. While an ERP component consistent with the P300 was visibly present in the data collected post-simulation, the test revealed no difference between the ERP response for oddball and distractor stimuli, t(12) = 1.92, p = 0.079. The ERP waveforms for pre- and post-simulation are presented in Figure 7.

Further, a paired t-test of P300 amplitudes pre- and post simulation confirmed a reduction in peaks post-simulation in comparison to pre, t(12) = 3.59, p = 0.0037, Cohen's d = 1.04 (see Figure 8). No correlation was found between fatigue scores and P300 amplitudes.

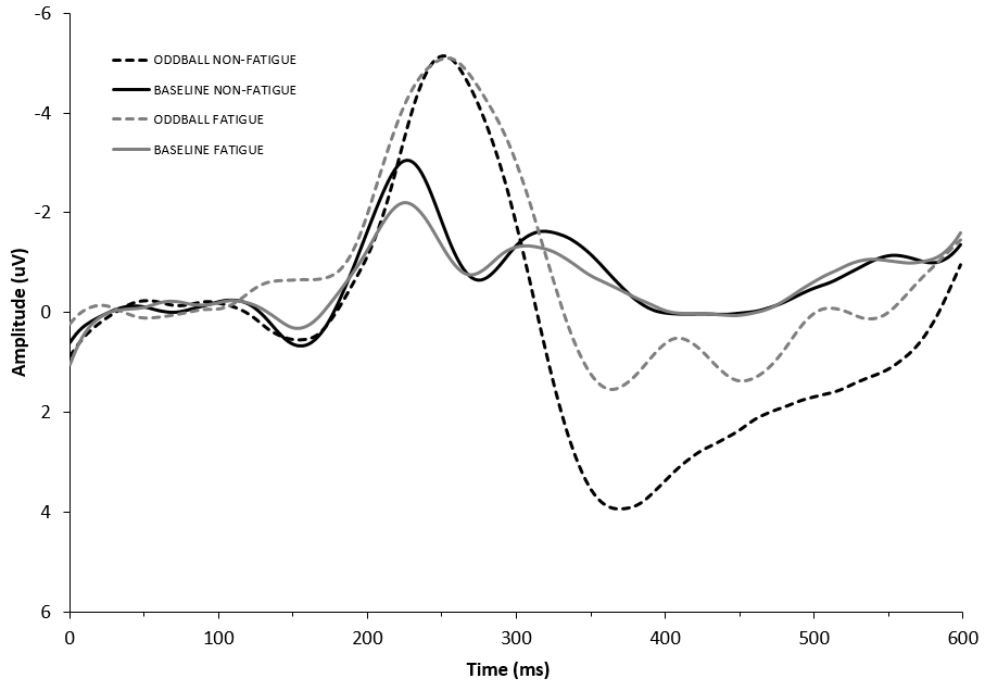


FIGURE 7. Conditional waveforms for distractor and oddball stimuli for pre medical simulation, overlaid on the waveforms for post-simulation in Experiment Three.

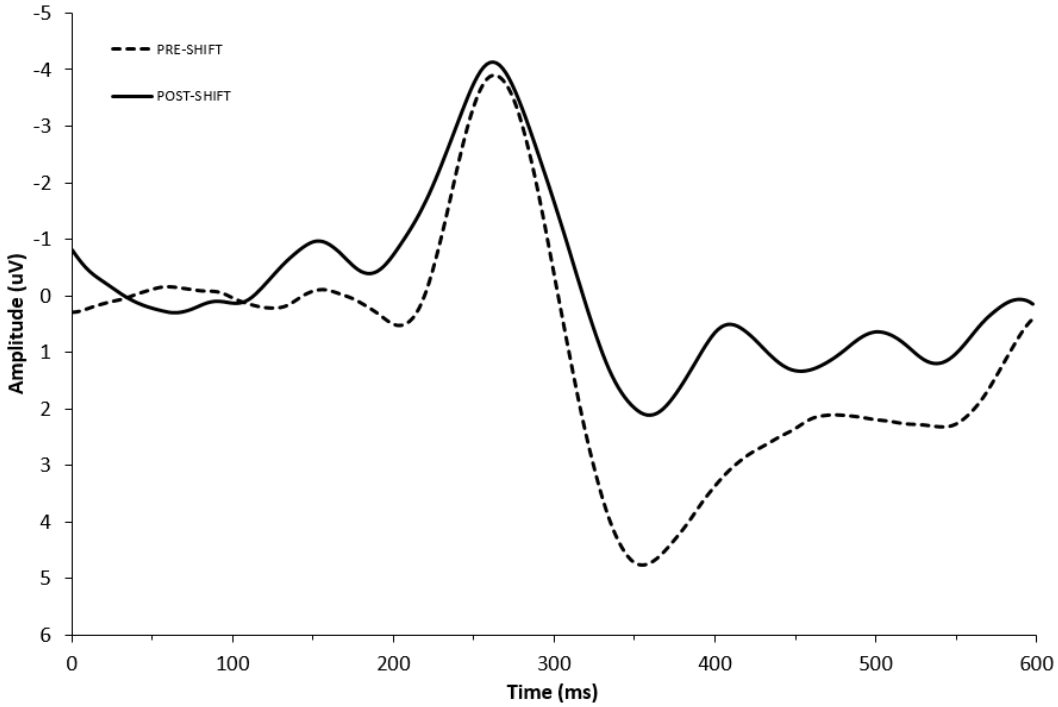


FIGURE 8. P300 difference waves computed as oddball-distractor for both pre and post- medical simulation in Experiment Three.

4.4 Summary

Experiment Three was first experiment in this study to be carried out in a medical environment, as well as the first in the study to include an experimental manipulation of fatigue. Thus, the purpose of this experiment was two fold: First, I aimed to successfully measure P300 amplitude using the MUSE EEG system and 3-minute oddball task, in a field-based medical setting, similar to the environment in which a physician might need to be assessed for fatigue. Second, I aimed to confirm the effect of fatigue on the P300 component, as supported by Experiments One and Two. As I expected, I was able to use the MUSE and associated oddball task to elicit the P300 component in the high-stress simulated medical environment, thus supporting the feasibility of using this tool in a hospital setting. In line with previous literature (Okuda 2009), analysis of the self-report data revealed that participants felt more mentally tired at the end of the experiment in comparison at the start, supporting the use of our within-subject experimental design. However, analysis revealed that there was no difference in scores on the M-OFER between the two time points, suggesting that participant's did not feel that their performance was impacted by the fatigue they were experiencing. This is not surprising giving previous studies that reported participants to have a limited ability to identify or assess their own fatigue-related impairments (Sahid et al,. 2010)

Analysis of the ERP data confirmed my hypothesis that the amplitude of the P300 would

be reduced post simulation in comparison to pre-, thus supporting a reduction in the P300 that is brought on by fatigue. Unlike Experiments One and Two however, I found no correlation between perceived fatigue score and peak amplitudes across subjects. This may be in part due to the small sample size of the study that resulted in fewer data points for the analysis. Visual inspection suggested the presence of the P300 component both pre and post simulation, however statistical testing suggested no difference in conditional waveforms, and thus no P300 component in the post-simulation dataset. Again, it is possible that due to the small sample size, power was not sufficient to detect small differences in the component in this group in comparison to all previous analyses. Alternatively, this outcome may be the result of severe fatigue developed during the duration of the simulation, indicating a cognitive deficit above and beyond what was present in previous experiments without experimental manipulation of fatigue. In sum, the outcome experiment 3 continues to support the MUSE assessment's sensitivity to fatigue given that the P300 was diminished post simulation in comparison to pre simulation. Further, while one subject was removed from analysis due to data quality, the overall success of data collection in Experiment Three supports the use of this procedure to collect fatigue-sensitive data on-site in a fast-paced medical environment.

CHAPTER 5: CONSIDERATIONS AND DISCUSSION

5.1 Implications of the Overall Study

The results of the above three experiments have confirmed my overall hypothesis that P300 component is sensitive to fatigue, and has potential for use as an objective fatigue assessment. Further, the outcome of experiments 2 and 3 confirmed that the sensitivity of the component to fatigue is intact when data is collected using a MUSE EEG headband, and further that usable data quantifying the P300 can be collected using a MUSE-oddball procedure in a field-based environment such as a hallway in a university campus, or in a busy hospital setting. Overall the outcome of this study supports the continued development of this tool to use as a measure and identification tool of mental fatigue in the medical workplace. Implementation of such a tool into the medical system would have broad impacts on the quality of care patients receive, as well as the quality of life experienced by medical practitioners.

While additional research and piloting will be required prior to recommending wide spread use of MUSE-oddball assessment be used to identify physician fatigue in a real-world environment, the success of the present study is enough to support the use of this technology for educational purposes. As discussed above, individuals appear to have limited skill in assessing the cognitive and performance deficits associated with their fatigue. Providing evidence of neural changes during the fatigue state, as demonstrated by the reduced P300 component during

high levels subject fatigue, may serve to educate the public on the dangers of fatigue and serve to increase the effort individuals devote to monitoring their own fatigue. For example, the methods employed in this study can be used to provide individuals with evidence as to the cognitive effects that exist during fatigue, and prolonged-wakefulness, beyond those they perceive. Additionally, upon repeated use, the MUSE assessment will be able to provide individuals with neural data of fatigue related changes in their own cognitive systems represented by the P300. For example, an individual who completes the MUSE assessment each day, may find that the magnitude of his or her P300 component is more, or even less sensitive to changes in schedule, or to long-duration shifts than the average individual. Thus, such practice would allow him or her to make better-educated decisions about their ability to work or complete particular tasks.

When evaluating personal fatigue, the recommendation from the Canadian Mental Protective Association is that doctors must consider the effects of fatigue on their thinking ability, and whether or not they are fit to provide patient care (Canadian Medical Protective Association, 2012). Unfortunately the competitive culture of medicine leads doctors to ignore the warning signs of fatigue and its associated cognitive deficits (Shirom, 2010; Kind, 2005). Physicians feel as though they are putting their patients first, and their own needs second, but this mentality leaves out the impact of the physicians fatigue on his or her patients (Wallace, Lemaire, & Ghali, 2009). Using the MUSE-EEG system to inform medical professionals of their

own fatigue can have the immediate effect of reducing the risks associated with relying on subjective feeling as the primary warning mechanism for fatigue-related performance deficits. Indeed, this method of fatigue management has been recommended by multiple reports, and by regulatory bodies in the field of occupational health and safety (College & Canada, 2013; Dawson & McCulloch, 2005; Elwood, 2007). Unfortunately, until the completion of this study, the potential for fatigue training through individualized on-site evaluation was limited due to the present lack of objective assessments (Egerton et al., 2015).

Long term, the incorporation of both the MUSE-based fatigue assessment, and MUSE-based individualized fatigue education into the medical profession will work to refocus the field on the need to develop, and carry out practical and effective methods of fatigue management. Indeed, a push for the movement of fatigue management away from the current system of prescriptive hours of service, and towards individualised monitoring has been consistent in medical profession over the previous decade (Canadian Medical Association, 2012, National Steering Committee for Resident Duty Hours, 2013).

While night work is unavoidable in the health care industry, research suggests some interventions are helpful in improving workers resiliency to shift-work (Occupational Cancer Research Center and the Institude for Work & Health, 2012). For example, regular, consistent schedules, promoting adaptation of circadian rhythm; controlled exposure to light and day; sleep

education and behavioural approaches directed towards physical activity, scheduling, and napping; and pharmacotherapy for promotion of sleep and wakefulness have all been shown to reduce the negative effects associated with shift-work based fatigue and performance deficits. Furthermore, in specific samples of medical professionals, scheduled naps during on-call and night work has been shown to improve cognitive skills, leading up to the following morning in comparison to a no-nap condition. Thus, while shift work may be unavoidable in the medical profession, the ability to assess acute fatigue in on-site physicians will provide the field with a warning mechanism for fatigue-related risk, as well overall reductions in fatigue if indications of the mental state are met with mitigating practices such as those described above.

5.2 Limitations and Future Directions

The primary limitation of the MUSE as a field-based assessment of fatigue is the quality of EEG data obtained from the MUSE. Indeed, while Krigolson and colleagues recently confirmed that this tool can be used to effectively collect ERP data for research purposes, they did discuss the issue of data loss, and in particular subject loss due to all experimental trials being removed artefact rejection. This is an unavoidable risk for any short-duration, limited-trial experimental paradigm but may be pose a bigger issue in the present study given the limited number of electrodes used, as well as the quality of the consumer grade MUSE EEG system. Indeed, data loss would be unfavourable in the situation of assessing time-pressured physicians

on-shift. However, in comparison to the loss of time and financial investment associated with EEG data loss using a standard research system, the economic implications of data loss using the MUSE in the field, is significantly reduced. In order to mitigate the risk of data loss, trained laboratory staff collected all data in experiments 2 & 3 and had previous experience with MUSE testing. Further, despite the short overall duration of the task, multiple signal checks will be conducted while EEG data are recording. Future studies aiming to provide support for the MUSE as a fatigue assessment in the medical, or alternative fields, should focus on improving data quality and reducing data loss through methodological or technological updates to the present study design.

With regards to study design itself the main limitation is the lack of performance measures collected from participants. The purpose of a on-site fatigue measure for physicians is to predict risks associated with deficits in cognitive and motor performance. Given the abundance of evidence supporting that the P300 predicts performance I chose not to add additional performance measures to this study at this time. However, given that the results of these experiments support the sensitivity of the P300 to fatigue, I recommend future studies further test this assessment by correlating P300 amplitudes with performance on decision-making task, such as the go-no-go paradigm. Advanced studies should aim to test the predictability of performance in medical and surgical tasks based on P300 components

magnitude, and in particular, ERP data collected with the MUSE. Such a design could be carried out using common laparoscopic simulation.

5.3 Ethical Considerations

There were several ethical concerns to review in preparation for this study. In particular, the potential risks to participants. Typically, EEG experiments pose few risks to participants. Commonly, these are limited to eye-strain from viewing a computer, or minor discomfort and itching caused by the placement of the electrode cap and associated gel. These risks applied to Experiment One in the proposed study. To reduce any associated discomfort of participants, they were introduced to the EEG-set up by a trained researcher who will address any questions they may have. Additionally, researchers completed multiple verbal check-ins with participants during set-up to confirm that they are comfortable. Given that the experimental task used throughout the study was substantially shorter in duration than other paradigms use to elicit the P300, eye strain was minimal.

Due to the use of the MUSE EEG system, risk of discomfort during the experiment was presumably reduced for Experiments Two and Three in comparison to one. However, due to the duration and involvement required from participants in Experiment Three, potential risks will include symptoms of fatigue that extend beyond the duration of the experiment. Additionally, participants may feel performance anxiety during the medical simulations. In order to reduce

potential anxiety, participants were given the opportunity to take part in an extended debriefing exercise with the simulation team. Given the purpose of the study, I did not propose any methods for reducing the risk of fatigue in participants. However, participants were reimbursed for the cost of a taxi to return home following the experiment as a measure of ensuring their safety. Given that all participants are enrolled in the medical residency program of their respective institutions, the anxiety and fatigue associated with this study should not have exceeded that which they experience in their regular work routines.

While containing more potential risk to participants, Experiment Three may have also offered the greatest benefit to those who take part. Research suggests that simulations are positive learning tools and that medical professionals experience less stress and anxiety transitioning into residency following work-related simulations such as this one (Laack, et al., 2010; Cohen et al., 2013; Rogers, et al., 2014). It was a possibility that subjects could feel more confident about their abilities and feel more prepared to begin their residencies following the completion of this study. The benefits of studies one, and two to participants are more limited, but include compensation and well as the opportunity to learn about fatigue and its associated risks from a trained researcher.

An additional ethical consideration associated with human research is the risk to participants' privacy and confidentiality. I employed methods throughout the full study in order

to ensure the privacy of participants and the confidentiality of all data collected. To begin, all data collected, both paper and digital, was be de-identified. Data were be labelled only with a code and not the participants' names or personal information. All documentation showing participants' names or personal information (i.e., consent forms) were be stored separately from all other data, in a secure filing area, in a secure laboratory. All digital data were stored on password-protected laboratory computers.

5.4 Summary

Due to the safety risks associated with physician fatigue, as well as the failure of fatigue management policies to mitigate these risks, a shift towards individual responsibility and monitoring of fatigue is required in the medical field (Dawson & McCulloch, 2005) Indeed, the Canadian Medical Association Code of Ethics states that doctors are ethically responsible to manage and assess their fatigue in order to mitigate any risks to patients. However, identifying one's own fatigue is not a straightforward process. As discussed in the previous sections of this proposal, research suggests that individuals underestimate, or are unable to detect, the effects of low levels of fatigue on their decision-making and judgment.

The three experiments completed in this study are the first steps in the development and validation of the MUSE-EEG system as a measurement for physician fatigue to be used in the field. The ultimate goal for the future is to use the MUSE system to improve fatigue management

in the medical profession, thus directly improving patient outcomes, and quality of life for physicians. Given the outcome the above experiments, I conclude that the MUSE-EEG assessment is a potential and promising method for fatigue assessment, that stands up to the cost effectiveness, objectivity, and portability required to be feasibly for use in medical field work.

In closing, the primary purpose this manuscript is to educate readers on the issues facing fatigue management in the medical field as well as the need for objective fatigue measures in order to improve these management systems. Finally, I aim to draw attention to the potential of the MUSE EEG system to be used as a portable, efficient, and effective tool of objective fatigue measurement in on-site physicians as well as the potential for future research to investigate its use as a training tool for fatigue identification. The further development of the MUSE EEG system for use as a fatigue assessment is paramount to improving fatigue management in the medical sector. Future implementation of objective assessments like the MUSE may limit the medical sector's reliance on one-size-fits-all POS regulations, and provide more individualised assessments and recommendations regarding physician fatigue. Fatigue information provided by the MUSE would provide an opportunity for tighter, more personalised management of fatigue in the medical workplace, contributing to reduced risk of fatigue-related medical mistakes, and better quality of patient care.

REFERENCES

Ackerman, P. L. (2011). Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications. Washington, D. C.: American Psychological Association.

Adell, E., Varhelyi, A., & Fontanta, M. D. (2011). The effects of driver assistance system for safe speed and distance - a real-life field study. Transportation Research Part C: Emerging Techologies , 19, 145-155.

Akerstedt, T. (2000). Consensus statement: Fatigue and accidents in transport operations. Journal of Sleep Research , 9, 395.

Akerstedt, T., Kecklund, G., Lowden, A., & Axelsson, J. (2000). Sleepiness and days of recovery. Transport Research Part F , 3, 251-261.

Aminoff, J. C., & Goodin, D. S. (2001). Long-latency cerebral event-related potentials in multiple sclerosis. Journal of Clinical Neurophysiology , 18, 372-377.

Antiel, R., Reed, D., Van Arenodonk, K., Wightman, S., Hall, D., & Porterfield, J. (2913). Effects of duty hour restrictions on core competencies, education, quality of life, and burnout among general surgery interns. (455, Ed.) Journal of the American Medical Association Surgery , 148 (5), 448.

Arnedt, J., Owens, J., & Crouch, M. (2005). Neurobehavioral performance of residents after

heavy night on call vs after alcohol ingestion. Journal of American Medical Association, 294 (9), 1345-1355.

Ayas, N., Barger, L., & Cade, B. e. (2006). Extended work duration and the risk of self-report percutaneous injuries in interns. Journal of the American Medical Association , 296 (9), 1055-1062.

Baker, A. (2000). Freight corp: Fatigue management program report. Adelaide, Australia: Centre for Sleep Research.

Baranski, J. V. (2007). Fatigue, sleep loss, and confidence in judgement. Journal of Experimental Psychology: Applied , 13 (4), 182196.

Barger, L., Ayas, N., Cade, B., Cronin, J., & Rosner, B. (2006). Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. PLoS Med , 3 (12), 2240-2248.

Barton, J., Spelten, E., Totterdell, P., Smith, L., & Folkard, S. (1995). Is there an optimum number of night shifts? Relationship between sleep, health & well-being. Work Stress , 9 (2), 109-123.

Beaumont, M., Batehat, D., Pierard, C., Coste, O., Doireau, P., & Van Beers, P. (2001). Slow release caffiene and prolonged (64-h) continuous wakefulness: effects on vigilance and cognitive performance. Journal of Sleep Research, 10 (4), 265-276.

Belz, S. M., Robinson, G. S., & Casali, J. G. (2004). Temporal separation and self rating of alertness as indicators of driver fatigue in commercial motor vehicle operations. Human Factors, 46 (1), 154-169.

Brown, I. (1994). Driver fatigue. Human Factors, 36, 298-314.

Burgess-Limerick, R., & Bowen-Rotsaert, D. (2002). Fatigue management program pilot evaluation: Phase 2 wave 3 report. Queensland, Australia: Global Institute of Learning and Development Consortium.

Canadian Medical Association. (2014). Management of Physician Fatigue. Canadian Medical Association Policy.

Canadian Medical Protective Association. (2013). CMPA Risk Fact Sheet: Patient Handover.

Canadian Medical Protective Association. (2012). The new realities of medical care.

CARE, R. A. (2009). Road Safety Evolution in EU. http://ec.europa.eu/transport/road safety/observatory/doc/historical evol.pdf.

Casanova-Gonzalez, M. F., Cabrera-Gomez, J. A., Aquino-Cias, J., Aneiros-Rivas, R., & Fernandez-Bermudez, R. (1999). Neurophysiological assessment in patients with clinically defined multiple sclerosis with special reference to P300 wave study. Revista de Neurologia, 29, 1134-1137.

Caterpillar Global Mining. (2008). Operator Fatigue Detection Technology Review. Peoria, IL:

Caterpillar.

Coenen, A. M. (1995). Neuronal activities underlying the electroencephalogram and evoked potentials of sleeping and waking. Neuroscience & Biobehavioral Reviews , 19 (3), 447-463.

Coplen, M., & Sussman, D. (2001). Fatigue and alertness in the United States railroad industry. Part II. Fatigue research in the office of research and development at the Federal Railroad Administration. US DOT, Volpe Centre: Operational and Safety Analysis Division.

Dawson, D., & Fletcher, A. (2001). A quantitative model of work-related fatigue: Background and definition. Ergonomics, 44 (2), 144-163.

Dawson, D., & McCulloch, K. (2005). Managing fatigue: It's about sleep. Sleep Medicine Reviews, 9 (5), 365-380.

Dawson, D., & Reid, K. (1997). Fatigue, alcohol and performance impairment. Nature, 388, 235.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., & Ott, G. E. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep, 20 (4), 267-277.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information

processing. In H. Begleite, Evoked Brain Potentials and Behavior (pp. 13-88). New York: Plenum Press.

Drolet, B., Sangisetty, S., Tracy, T., & Cioffi, W. (2013). Surgical residents' perceptions of 2011 Accreditation Council for Graduate Medical Education duty hour regulations. Journal of American Medical Association Surgery, 148 (5), 427-433.

Eastridge, B., Hamilton, E., O'Keefe, G., Rege, R., & Vaentine, R. e. (2003). Effect of sleep deprivation on the performance of laparoscopic surgical skill. The American Journal of Surgery, 186, 169-174.

Elger, T., Bethke, F., Frese, A., Luettmann, R. J., Buchheister, A., Ringelstein, E. B., et al. (2002). Event-related potentials in difference subtypes of multiple sclerosis-a cross-sectional study. Journal of Neurological Science, 205, 35-40.

Folkard, S. (1997). Black times: Temporal determinants of transport safety. Accidents Analysis & Prevention, 29 (4), 417-430.

Folkard, S., & Tucker, P. (2003). Shift work, safety and productivity. Occupational Medicine, 53, 95-101.

Gander, P. H., Merry, A., Millar, M. M., & Weller, J. (2000). Hours of work and fatigue-related error: A survey of New Zealand anaesthetists. Anaesthesia & Intensive Care Medicine, 25 (2), 178-183.

Gander, P. H., Waite, D., McKay, A., Seal, T., & Millar, M. M. (1998). An integrated fatigue management programme for tanker drivers. In L. Hartley, Managing fatigue in transportation: Proceedings of the third fatigue in transportation conference. Freemantle, Western Australia: Pergamon.

Gawron, V. J., French, J., Funke, D., Hancock, P. A., & Desmond, P. A. (2001). An overview of fatigue. Stress, workload, and fatigue, 581-595.

Grandjean, E. (1979). Fatigue in industry. British Journal of Industrial Medicine, 36, 175-186.

Grandjean, E. (1988). Fitting the Task to the Man. A Textbook of Occupational Ergonomics. London: Taylor & Francis.

Haines, V., Marchand, A., Rousseau, V., & Demers, A. (2008). The mediating role of work-to-family conflict in the relationship between shiftwork and depression. Work and Stress, 22 (4), 341-356.

Haraldsson, P. O., & Akerstedt, T. (2001). Sleep disturbances - Greater traffic hazard than alcohool. Causes, risks and treatment. Lakartidningen , 98 (25), 3018-3023.

Harmony, T., Fernandez, T., Silva, J., Bernal, J., Diaz-Comas, L., Reyes, A., et al. (1996). EEG delta activity: An indicator of attention to internal processing during performance of mental tasks. International Journal of Physiology, 24, 161-171.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: a

review. Journal of Experimental Psychology: Applied, 6 (3), 236-249.

Hartley, L. R. (1995). Fatigue and Driving. Bristol: Taylor & Francis.

Hartley, L. R., Arnold, P. K., Smythe, G., & Hansen, J. (1994). Indicator of fatigue in truck drivers. Applied Ergonomics, 25, 143-156.

Horne, J. A., & Baulk, S. D. (2004). Awareness of sleepiness when driving. Psychophysiology, 4, 97-110.

Horne, J. A., & Reyner, L. A. (1995a). Driver Sleepiness. Journal of Sleep Research, 4 (2), 23-29.

Horne, J., & Reyner, L. (1995a). Sleep related vehicle accidents. BMJ, 310, 565-567.

Horvath, M., Frantik, E., Kopriva, K., & Meissner, J. (1976). EEG theta activity increase coinciding with performance decrement in a monotonous task. Activas Nervosa Superiios (Pragye) , 18 (3), 207-210.

Howard, S., Gaba, D., & Smoth, B. (2003). Simulation study of rested versus sleep-deprived anesthesiologists. Anesthesiology, 98, 1345-1355.

Idogawa, K. (1991). On the brain wave activity of professional drivers during monotonous work. Behavior metrika, 30, 23-34.

Institutes BC. (1999). Occupational health and safety assessment series (OHSAS 18001): Occupational health and safety management systems-specification. London: British

Standards Institute.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980a). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychology, 17, 259-273.

Isreal, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. (1980b). The event-related potential as an index of display monitoring workload. Human Factors, 22, 211-224.

Jamal, M. (2004). Burnout, stress and health of employees on non-standard work schedules: a study of Canadian workers. Stress and Health, 20, 113-119.

Johns, M. W., Chapman, R., Crowely, K., & Tucker, A. (2008a). A new method for assessing the risks of drowsiness while driving. Somnologie, 12, 66-74.

Johns, M. W., Tucker, A., Chapman, R., Crowley, K., & Michael, N. (2007). Monitory eye and eyelid movements by infrared reflectance occulography to measures drowsiness in drivers. Somnologie-Schlafforschung Schlafmedizin, 11, 234–242.

Johnson, D. E., & Donchin, E. (1980). P300 and stimulus categorization: two plus one is not so different from one plus one. Psychophysiology, 17, 167-178.

Khardi, S., & Vallet, M. (1994). Drivers vigilance, analysis of differences in vigilance states assessments by physiological and mechanical indicators. Telematics for Transport, (pp. 1991-1998). Paris.

Knutsson, A., & Boggild, H. (2010). Gastrointestinal disorders among shift workers. Scandinavian Journal of Work and Environmental Health, 36 (2), 85-95.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1983). An analysis of the processing demands of a complex perceptual-motor task. Human Factors, 25, 597-622.

Krueger, K., & Halperin, E. (2010). Perspective: Paying physicians to be on call: A challenge for academic medicine. Academic Medicine, 85 (12), 1840-1844.

Krupp, L. B., Larocca, N. G., Muir-Nash, J., & Steinberg, A. D. (1989). The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Archives of Neurology, 46, 1121-1123.

Lal, S. K., & Craig, A. (2000b). Psychophysiological effects associated with drowsiness: Driver fatigue and electroencephalography. International Journal of Psychophysiology, 35, 39.

Landrigan, C., Rothschild, J., & Cronin, J. (2004). Effect of reducing Intern's work hours on serious medical errors in intensive care units. New England Journal of Medicine, 351, 1838-1848.

Leape, L. L., Brennan, T. A., Laaird, N., Lawthers, A. G., Logalio, A. R., & Barns, B. A. (1991). The nature of adverse events in hospitalized patients. New England Journal of Medicine, 327 (6), 377-384.

Leiberman, H. R., Tharion, W. J., Shukitt-Hale, B., Speckman, K. L., & Tulley, R. (2002).

Effects of caffeine, sleep loss, and stress on cognitive performance and mood during U.S. Navy SEAL training. Sea-air-land. Psychopharmacology (Berl.), 164 (2), 250-261.

Lenne, M. G., Triggs, T. J., & Redman, J. R. (1997). Time of day variations in driving performance. Accident Analysis & Prevention, 29 (4), 431-437.

Lisper, H. O., Laurell, H., & Van Loon, J. (1986). Relation between time falling asleep behind the wheel on a closed track and changes in subsidiary reaction time during prolonged driving on a motorway. Ergonomics, 29 (3), 445-453.

Lockley, S., Barger, L., Ayas, N., Rothschild, J., & Czeisler, C. (2007). Effects of health care provider work hours and sleep deprivation on safety performance. The Joint Commission Journal on Quality and Patient Safety, 3 (11), 7-18.

Lowden, A., Kecklund, G., Axelsson, J., & Akerstedt, T. (1998). Change from an 8-hour shift to a 12-hour shift, attitudes, sleep, sleepiness and performance. Scandinavian Journal or Work, Environmental & Health, 24 (3), 69-75.

Magliero, A., Bashore, R. B., Coles, M. G., & Donchin, E. (1984). On dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 21, 171-186.

Mahon, G., & Cross, T. (1999). The fatigue management program: Alternatives to prescription. Queensland, Australia: Queensland Transport.

Mangun, G. R., & Hillyard, S. A. (1987). The spatial allocation of visual attention as indexed by

event-related brain potentials. Human Factors, 29, 195-211.

Mascord, D. J., & Heath, R. A. (1992). Behavioural and physiological indices of fatigue in a visual tracking task. Journal of Safety Research, 23, 19-25.

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transportation Research Part F: Traffic Psychology and Behaviour, 12 (3), 218-224.

Maycock, G. (1997). Sleepiness and driving: The experience of UK car drivers. Accidents Analysis & Prevention, 29 (4), 453-462.

McCulloch, K., Fletcher, A., & Dawson, D. (2003). Moving toward a non-prescriptive approach to fatigue management in Australian aviation: A field validation. Canberra, Australia: Civil Aviation Safety Authority.

Moller, H. J., Kayumov, L., Bulmash, E. L., Nhan, J., & Shapiro, C. M. (2006). Simulator performance, microsleep episodes, and subjective sleepiness: normative data using convergent methodologies to assess driver drowsiness. Journal of Psychosomatic Research, 61, 335-342.

Moonesinghe, S., Lowery, J., Shahi, N., Millen, A., & Beard, L. (2011). Impact of reduction in working hours for doctors in training on postgraduate medical education and patients' outcomes: Systematic review. BMJ, 342.

Murata, A. (1994). Experimental discussion on measurement of mental workload-evaluation of mental workload by HRV measures. EICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E77-A, 409-416.

Murata, A., & Iwase, H. (2000). Evaluation of mental workload by variability of pupil area. IEICE Transaction on Information and Systems E83-D, 1187-1190.

Neri, D. F., Shappell, S. A., & DeJohn, C. A. (1992). Simulated sustained flight operations and performance. Part 1. Effects of Fatigue. Military Psychology, 4 (3), 137-155.

Neuman, U., Ullsperger, P., Gille, H. G., & Erdman, U. (1986). Effects of graduated processing difficulty on P300 component of the event-related potential. Zeitschrift fur Psychology, 194, 25-37.

Nicol, A., & Botterill, J. (2004). On call work and health: A review. Environmental Health, 3, 1-11.

Nordbakke, S., & Sagberg, F. (2007). Sleepy at the wheel: Knowledge, symptoms and behaviour among car drivers. Transportation Research Part F, 10, 1-10.

Occupational Cancer Research Center and the Institute for Work & Health. (2012). Can the health effects of shift work be mitigated? A summary of select interventions.

(1990). Occupational Health and Safety Act, in R. S. O. 1990.

Okogbaa, O. G., Shell, R. L., & Filipusic, D. (1994). On the investigation of the

neruophysiological correlates of knowledge worker mental fatigue using EEG signal. Applied Ergonomics, 25, 355-365.

Parks, D., Yetmen, R., McNeese, M., Burau, K., & Smolensky, M. (2000). Day-night pattern in accidental exposures to blood-borne pathogens among medical students and residents. Chronobiology International, 17 (1), 61-70.

Philbert, I. (2005). Sleep loss in residents and physicians: A meta-analytic examination. Sleep, 28, 1392-1402.

Philbert, I., Nasca, T., Brigham, T., & Shapiro, J. (2007). Duty-hour limits and patient care and residence outcomes: Can high-quality studies offer insight into complex relationships? Annual Review of Medicine, 54, 467-483.

Phillip, P., Sagaspe, P., Taillard, J., Guilleminault, C., Sanchez-Ortuno, M., Akerstedt, T., et al. (2003). Fatigue, sleep restriction, and performance in automobile drivers: A controlled study in natural environment. Sleep, 23 (3), 277-280.

Phillip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Akerstedt, T., et al. (2005). Fatigue, sleepiness, and performance in simulated vs real driving conditions. Sleep, 28 (12), 1151-1516.

Phillp, Ghorayeb, I., Leger, D., Menny, J. C., Bioulac, B., Dabadie, P., et al. (1997). Objective measurement of sleepiness in summer vacation long-distance drivers. EEG Journal of

Clinical Neurophysiology, 102, 383-389.

Piras, M. R., Magnano, I., Canu, E. D., Paulus, K. S., SAtta, W. M., Soddu, A., et al. (2003). Longitudinal study of cognitive dysfunction in multiple sclerosis: Neuropsychological, neurological, and neurophysiological findings. Journal of Neurology, Neurosurgery, and Psychiatry, 74, 878-885.

Pokryszko-Dragan, A., Zagrajek, M., Slotwinski, K., Gruszka, E., Bilinska, M., & Podemski, R. (2009). Neuropsychological testing and event-related potentials in the assessment of cognitive performance in the patients with multiple sclerosis-a pilot study. Clinical Neurology and Neurosurgery, 111, 503-506.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. Clinical Neurophysiology, 118 (10), 2128-2148.

Polich, J., Romine, J. S., Sipe, J. C., Aung, M., & Dalessio, D. J. (1992). P300 in multiple sclerosis: A preliminary report. International Journal of Psychophysiology, 12, 155-163.

Queensland Transport. (2001). Fatigue management program: Discussion paper. Queensland Transport.

Rhodes, W., & Gil, V. (2002). Development of a fatigue management system program for Canadian marine pilots. Montreal: Transportation Development Centre.

Rothschild, J., Koehane, C., & Rogers, S. (2009). Risks of complications by attending physicians

after performing nighttime procedures. JAMA, 302, 1564-1572.

Schernhammer, E., Laden, F., Speizer, F., Willet, W., Hunter, D., & Kawachi, I. e. (2003). Night-shift work and risk of colorectal cancer in the Nurses' Health Study. Journal of the National Cancer Institute, 95 (11), 825-828.

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgment of vigilance state during prolonged monotonous daytime driving. Accidents Analysis & Prevention, 41, 1087-1093.

Schroeder, D. J., Rosa, R., & Witt, L. A. (1998). Some effects of 8-vs 10-hour work schedules on the test performance/ alertness of air traffic control specialists. International Journal of Industrial Ergonomics, 21, 307-321.

Schubert, M., Johannes, S., Koch, M., Wieringa, B. M., Dengler, R., & Munte, T. F. (1998). Differential effects of two motor tasks on ERPs in an auditory classification task - evidence of shared cognitive resources. Neuroscience Research, 30 (2), 125-134.

Shields, M. (2002). Shift work and health. Health Reports, 13 (4), 11-33.

Smith, S. S., Horswill, M. S., Chambers, B., & Wetton, M. (2009). Hazard perception in novice and experienced drivers: the effects of sleepiness. Accident Analysis & Prevention, 41 (4), 729-733.

Standards Australia. (2001). The Australian/New Zealand standard for occupational health and

safety management systems (AS4801). Sydney; New South Wales: Council of Standards Australia and New Zealand.

Sundgren, M., Nikulin, V. V., Maurex, L., Wahlin, L., Piehl, F., & Brismar, T. (2015). P300 amplitude and response speed related to preserved cognitive function in relapsing-remitting multiple sclerosis. Clinical Neurophysiology, 126, 689-697.

Taffinder, N., McManus, I., Hul, Y., Russell, R., & Darzi, A. (1998). Effect of sleep deprivation on surgeon's dexterity on laparoscopy simulator. The Lancet, 352, 1191.

The Parliment of The Commonwealth of Australia. (2000). Beyond the Midnight Oil: An Inquiry into Manage Fatigue in Transport. Australia: House of Representatives Standing Committee on Communication, Transport, and the Arts.

Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: A simulator study. Accidents Analysis & Prevention, 35, 381-391.

Uetake, A., & Murata, A. (2000). Assessment of mental fatigue during VDT task using event-related potential (P300). Proceedings of the 2000 IEEE International Workshop on Robot and Human Interactive Communication, (pp. 235-240).

Ullsperger, P., Metz, A. M., & Gille, H. G. (1988). The P300 component of the event-related brain potential and mental effort. Ergonomics, 31, 1127-1137.

Ullsperger, P., Neumann, U., Gille, H. G., & Pietschann, M. (1986). P300 component of the ERP

as an index of processing difficulty. In F. Flix, & H. Hagendor, Human Memory and Cognitive Capabilities (pp. 723-731). Amsterdam: North-Holland.

Ulmer, C., Wolman, D., & Johns, M. (2008). Resident Duty Hours: Enhancing Sleep. Institute of Medicine, Committee on Optimizing Graduate Medical Trainee (Resident) Hours and Work Schedule to Improve Patient Safety. Washington, DC: National Academic Press.

Van Dongen, H. P., Baynard, M. D., Maislin, G., & Dinges, D. F. (2004). Systematic individual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep, 27 (3), 423-433.

Van Dongen, H., Baynard, M., & Maislin, G. (2004). Systematic individual difference in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep, 27, 423-433.

Vlopp, K., Rosen, A., & Rosenbaum, P. P. (2007). Mortality among hospitalized Medicare beneficiaries in the in the first 2 years following ACGME resident duty hour reform. JAMA, 298, 975-983.

Volpp, K., Rosen, A., & Rosenbaum, P. (2997). Mortality among patients in VA hospitals in the first 2 years following ACGME resident duty hour reform. JAMA, 298 (9), 984-992.

Vyas, M., Garg, A., Iansavichus, A., Costella, J., Donner, A., & Laugsand, L. e. Shift work and vascular events: Systematic review and meta-analysis. British Medical Journal, 345.

Wallace, J., Lemaire, J., & Ghali, W. (2009). Physician wellness; a missing quality indicator. The Lancet, 374 (9702), 1714-1724.

Wickens, C. D. (1979). Measures of workload, stress and secondary tasks. In N. Moray, Mental Workload: Its Theory and Measurement (pp. 79-99). New York: Plenum Press.

Williamson, A. M., Feyer, A., & Friswell, R. (1996). The impact of work practices on fatigue in long distance truck drivers. Accident Analysis & Prevention, 28 (6), 709-719.

Williamson, A., & Freyer, A. (2000). Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol inoxication. Occupational and Environmental Medicine, 57, 649-655.

Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. Accident Analysis & Prevention, 43 (2), 498-515.

APPENDIX A

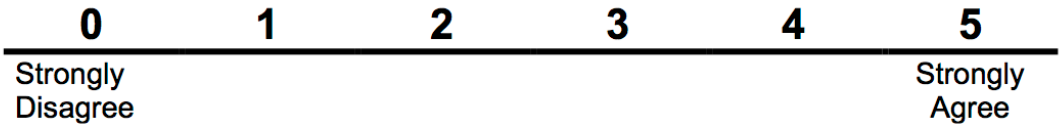
Example of Fatigue Measure used in Present Study

Researcher's Verbal Instructions:

"Please use this scale to let us know how mentally tired you feel right now, in this very moment."

Use the line below to report how the following statement applies to you right now.

"I feel mentally tired"



APPENDIX B

Example of Modified - Occupational Fatigue Exhaustion Recovery Scale

Scale	Strongly disagree	Disagree	Slightly disagree	Neither agree or disagree	Slightly agree	Agree	Strongly agree
1. I often feel I'm 'at the end of my rope'	0	1	2	3	4	5	6
2. I often dread waking up to another day of my work.	0	1	2	3	4	5	6
3. I often wonder how long I can keep going at my work.	0	1	2	3	4	5	6
4. I feel that most of the time I'm just "living to work."	0	1	2	3	4	5	6
5. Too much is expected of me in my work	0	1	2	3	4	5	6
6. After a typical work period I have little energy left	0	1	2	3	4	5	6
7. I usually feel exhausted when I get home from work	0	1	2	3	4	5	6
8. My work drains my energy completely everyday	0	1	2	3	4	5	6
9. I usually have lots of energy to give my family or friends.	0	1	2	3	4	5	6
10. I usually have plenty of energy left for my hobbies and other activities after I finish work	0	1	2	3	4	5	6
11. I never have enough time between work shifts to recover my energy completely	0	1	2	3	4	5	6
12. Even if I am tired from one shift, I am usually refreshed by the start of the next shift.	0	1	2	3	4	5	6
13. I rarely recover my strength fully between work shifts	0	1	2	3	4	5	6
14. Recovering from work fatigue between work shifts isn't a problem for me	0	1	2	3	4	5	6
15. I'm often still feeling fatigued from one shift by the end of the time I start the next one.	0	1	2	3	4	5	6