Cardiac Function, Pupillometry, Subjective Workload and Performance of Commercial Pilots During Jet Airplane Handling Manoeuvres

Samuel Clément-Coulson

A Thesis

In

The Department

Of

Psychology

Presented in Partial Fulfillment of the Requirements for the Degree of Master of Arts

(Psychology) at Concordia University

Montreal, Quebec, Canada

August 2020

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Samuel Clément-Coulson

Entitled: Cardiac Function, Pupillometry, Subjective Workload and Performance of Commercial Pilots During Jet Airplane Handling Manoeuvres

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Psychology, Research)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

____Andreas Arvanitogiannis__ Chair

____Emily Coffey____________ Examiner

____Jennifer McGrath________ Examiner

____Aaron Johnson__________ Supervisor

Approved by: _____Aaron Johnson_____________

Chair of Department or Graduate Program Director

_____Pascale Sicotte______________

Dean of Faculty

_____02 September 2020__________

Date

Abstract

Cardiac Function, Pupillometry, Subjective Workload and Performance of Commercial Pilots During Jet Airplane Handling Manoeuvres

Samuel Clément-Coulson

Projected demand for commercial pilots is expected to exceed current training programs. As such, next-generation training programs are focused on developing competencies through practice sessions targeting specific manoeuvres. These sessions are most effective when conducted near a trainee's maximum workload. Thus, measuring and understanding the relationship between workload and performance in the cockpit is of prime importance. Currently, workload measurement in the cockpit relies on subjective measures. More recently, heart rate, heart rate variability and pupillometry have shown promise as objective, continuous indicators of workload. We investigated relationships between these psychophysiological measures, NASA-TLX subjective workload ratings and aircraft handling performance in commercial pilots flying a jet airplane simulator. We found that (1) heart rate and heart rate variability patterns were stable between participants, showing clear differences between manoeuvre types, but not manoeuvre difficulty, (2) pupil diameter is not an effective measure of workload in the simulator cockpit, likely due to the differences in brightness, and (3) psychophysiological measures of workload can supplement subjective workload reports to better understand performance.

Acknowledgements

Dr. Aaron Johnson, thank you for supporting me ever since we first met during my undergraduate degree. There have been many challenges and twists between the start of our aviation research and the completion of this project, and you have been there at every stage of the way. You have ignited a passion for science. Your support and instruction have been invaluable, enabling academic, professional and personal growth. Thank you.

Concordia Vision Lab members Alaa Boutella, Ramiya Veluppillai, Karine Elalouf and Pia Karpowitz, thank for your crucial help throughout this project. Antoine Carrière, Paul Vadja, Caro Rohleder and Leon Franzen, thank you for the endless airplane talks and discussions!

Mom, Dad, thank you for supporting my passion for aviation and science. Jock and Renée, thank you for teaching me the importance of hard work and dedication. Gilles et Rollande, merci de m’avoir enseigné que tout est possible lorsqu’on y croit.

J’aimerais remercier le Centre Québécois de Formation Aéronautique pour le support depuis mes débuts au CQFA en 2012. Au-delà d’avoir été essentiel à ce projet de recherche, avec plusieurs discussion et l’accès au simulateur, le CQFA est une école qui m’a aidé à devenir un professionnel de l’aéronautique, à travers d’innombrables apprentissages personnels et techniques. En particulier, j’aimerais remercier Jean Laroche pour son aide pendant le projet, ainsi que David Gagnon, Joël Avenel, Martial Taureau et Gilles Traversy pour leur enseignement dans mes débuts en aviation.

À tous les pilotes qui ont donné leur temps en participant dans ce projet, ou qui ont discuté de leurs expériences et idées, merci. Ces discussions me passionnent et me motivent à faire de la recherche pour l’aviation. Dans ces temps difficiles pour notre industrie, c’est cette passion contagieuse qui me pousse à continuer. #UnitedByWings

CONTRIBUTION OF AUTHORS

All sections were written with general feedback from my supervisor, Dr. Aaron Johnson. Several colleagues contributed to the collection of data: Alaa Boutella, Pia Karpowitz and Karine Elalouf. Ramiya Veluppillai assisted in data cleaning and preparation, as well as graph formatting.

Table of Contents

List of Tables

List of Figures

Introduction

Since World War II, the increased requirements for competent aviation professionals and adequate equipment have led to a new area of applied research: human factors (Waterson, 2011). Human factors is the branch of science that investigates the “relation between people and the machines they operate through the systematic application of human sciences integrated within the framework of systems engineering” (Jensen, 1997). In the aviation industry, research into human factors aims to improve our understanding of the physiological and psychological limitations of human operators, as a critical component in the operation of an aircraft. In aviation, human factors researchers have focused on two complementary strategies: designing systems to maximize the effectiveness of the human-machine combination as a unit, and improving the training delivered to individuals involved in operations. In aviation human factors, training refers to selecting and training operators until they become experts on the systems they operate (Jensen, 1997). However, the International Civil Aviation Organization, the United Nation’s specialized agency, forecasts a demand for 620 000 new pilots by 2036, most of which will be new pilots yet to be trained (ICAO, 2017). This demand for new pilots is expected to impose significant demands on flight training resources. Further complicating the situation, recent industry trends show decreasing numbers of flight instructor qualification holders and lower experience among qualified instructors as most quickly progress to airline positions (Standing Committee on Transport Infrastructure and Communities, 2019). Following an investigation into the current and projected states of the aviation industry, the Canadian House of Commons proposed 10 recommendations to support flight training. Three of these recommendations support the development of technologies and their application in training pilots to attain required competency. Specifically, the committee recommended that training methods focus on pilot competence development, that the Government of Canada support the development of new technologies for flight training (i.e., flight simulation equipment), and that regulations be modernized to allow implementation of these new technologies (Standing Committee on Transport Infrastructure and Communities, 2019). In line with these objectives, the present thesis investigates the use of flight simulator data, psychophysiological, and self-report measures in assessing pilot workload, which could enable efficient use of training resources by personalizing pilot training to individual pilots' training needs.

Expert Performance in Aviation

Researchers in the field of cognitive science have developed a number of theoretical frameworks that allow us to study the nature of expertise. For example, the information processing model of human problem solving proposes that cognition relies on physical structures such as sensory organs and other brain structures; and mental processes developed through learning and experience to perceive and process information from one's environment and act upon it (Newell & Simon, 1972). This means that expert performance is the product of learning through practice, not innate ability. Hence, experts differ from the general population by having superior domain-specific knowledge and mental processes allowing them perform better (Ericsson et al., 1993). Specifically, experts distinguish themselves from novice performers by having more elaborate understanding of how their environment works, and how to function within an environment (Ericsson & Charness, 1994). This is known as an expert's mental models, or mental representations, of systems, tasks and situations. In comparison to novices, experts build mental representations of a situation much quicker, require less information, and show shorter decision times (Ericsson & Charness, 1994; Kennedy et al., 2019; Schriver et al., 2008).

A defining characteristic of expertise is that it is a highly domain-specific attribute built over extensive periods of deliberate practice, conscious behaviour aimed at improving one's skill (Ericsson et al., 1993). For example, skills acquired by an expert airplane pilot do mean the pilot is an expert in automobile racing. The pilot would have to conduct deliberate practice specific to automobile racing in order to develop expertise in that domain. In order to maximize improvement, and qualify as deliberate practice, four conditions must be met: well-defined goals, full concentration, immediate feedback, and repetition of exercises with gradual refinements of technique (Ericsson, 2008; Ericsson et al., 1993). Through deliberate practice, individuals build mental models of the skill, which include characteristics of the objects in their environment, understanding of the interactions between objects in environment, and a person's ability to interact with these objects and their environment (Endsley, 2000). Mental models allow experts to overcome human perception limitations and limited information processing speed by relying on learned information and processes (Newell et al., 1958). Specifically, experts classify and match situations with stored representations and recall an appropriate course of action (Orasanu & Fischer, 1997). This matching reduces cognitive load by eliminating the need for conscious processing of all elements (Schriver et al., 2008) . Better mental models have been associated with superior flying performance, as defined by increased flying accuracy, understanding of aircraft states, better decision-making, and shorter response time to abnormal situations (Endsley, 2006). Hence, deliberate practice of exercises, or manoeuvres, that develop trainees' mental models of successful flying practices must be a core component of flight training.

Training for Expertise

To develop a mental model in training, early studies of performance have demonstrated a link between time spent doing an activity and one's mental representations of that activity (Ericsson et al., 2006). In aviation, total flying experience is traditionally accepted by most training programs and pilot certification assessment as a measure of one's skill. At first glance, this make sense and is backed up by accident rate data. For example, as pilots accumulate more flying time, they are less likely to enter into an accident caused by human error (Bazargan & Guzhva, 2011). This makes sense, as total time spent doing a task is generally linked to one's level of expertise (Ericsson & Charness, 1994). Findings such as these support measurement of expertise based on total experience, and appear to support using total flying time as an industry-accepted measure of a pilot's expertise. However, estimates of expertise based on total experience are correlational, omitting a causal link between experience and expertise. Specifically, is there a specific behaviour that increases one's expertise?

Ericsson's theory of deliberate practice proposes that simple repetition of a task is not enough to improve skill. Instead, one must engage in a rigorous type of practice called deliberate practice (Ericsson et al., 1993). Building on the well-known concept of practice, involving repetition of a task to improve skill, deliberate practice adds on four criteria that differentiate it and make it the most efficient way of improving one's skill. It requires a clearly defined performance objective, full concentration and engagement in the task, refining technique through repetition of the exercise and immediate, formative feedback. Flight training curriculums and protocols explicitly include these four components of deliberate practice

(Transport Canada, 2004). Specifically, lessons are designed to be slightly outside a candidates' existing skill level with progressively complex exercises. Deliberate practice also requires feedback from an instructor, where trainees receive feedback to improve performance and receive instructions for future exercises. Thus, debriefings immediately after exercises and training sessions are conducted as part of flight training. Due to all the components aimed at improvement of skill, time spent conducting deliberate practice predicts skill learning much more than any other type of activity (Duckworth et al., 2011; Ericsson et al., 1993; Kellogg & Whiteford, 2009). An implication of these findings is that task repetition day-to-day operations contributes to ones' total flight time, but contributes little towards improving the pilot's skills. Improving skills occurs when one engages in deliberate practice. One engages in deliberate practice during training to obtain or renew qualifications. Thus, pilot expertise is highly related to qualification level. For example, while investigating unauthorized entry into active runways by pilots, Van Benthem and Herdman (2013) found that pilots with higher qualifications (i.e., instrument flying rating, commercial pilot, military qualifications) had reduced tendencies of committing mistakes such as runway incursions. The obtention of higher qualifications in aviation is linked to spending more time in training, hence more deliberate practice. In commercial aviation, pilots generally work in pairs where tasks are shared between crew members. Even at the crew-level, deliberate practice was shown to significantly improve flight crews' decision-making during emergencies (McKinney & Davis, 2003).

Together, the findings of McKinney and Davis (2003) and Van Benthem and Herdman (2013) suggest that deliberate practice is part of the curriculum for aviation license upgrades, while accumulating "flight hours" conducting regular flights may not lead to an increase in a pilot's skill. Todd & Thomas (2012) found no performance differences in technical skills (e.g., meeting approach performance criteria, risk management) between first officers with low total flight time (< 1500 hours) and their more experienced counterparts (> 1500 hours). These findings further support the idea that lack of opportunities for deliberate practice outside of qualification-upgrade courses effectively limits a pilot's skill progression. Limited opportunities for deliberate practice lead to well-trained and well-intentioned workers' expertise plateau due to a lack of opportunities for deliberate practice after entering the workforce. From a longitudinal perspective, published flight training curriculums substantiate this idea. During a pilot's early career, training practices continuously expose pilots to deliberate practice. During the obtention of their private pilot's license, pilots are exposed to a series of exercises with gradual increases in difficulty aimed at helping them master technical skills of flying an airplane. During their commercial training, the same exercises are repeated, but performance criteria are heightened, and feedback remains mandatory. In their early career, they must master these exercises on gradually more complex aircraft, and senior pilots are available to provide feedback. However, once established in a position, commercial pilots' recurrent training no longer meets the criteria of deliberate practice.

Commercial pilots in recurrent training are exposed to fixed performance objectives (Transport Canada, 2017), and feedback from instructors is limited. For example, feedback on workload management from instructors refers only to observable aspects of the performance. Instructors in commercial aviation are not required to have the same teaching-oriented training as those who teach private and commercial flying. In order to take full advantage of mandated recurrent training for commercial pilots, training sessions should be personalized to fulfill the criteria of deliberate practice. Additionally, formative feedback would have to be provided during training sessions. The nature of piloting tasks poses significant challenges in implementing these two requirements of deliberate practice. This occurs because optimal performance in aviation is difficult to quantify, and instructors in commercial aviation have limited insight into the internal processes of the individuals they are evaluation as they perform training routines in flight simulators due to the environment's physical limitations.

Quantifying and measuring

Deliberate practice relies on accurate observations of a trainee's performance to provide formative feedback. Chuck Yeager, the aviator who first broke the sound barrier, once said "If you can walk away from a landing, it's a good landing. If you use the airplane the next day, it's an outstanding landing." If evaluation of aviation performance was based on repeated measures of a categorical outcome, such as a successful landing, this would be an adequate definition. However, performance in aviation cannot rely on categorical outcomes because errors in flying activities are potentially fatal. One must always have a successful landing in order to be able to repeat the exercise and continue practicing. Therefore, assessment of performance in piloting focuses heavily on deviation from ideal techniques and procedures (Transport Canada, 2017). Performance on technical skills is relatively easy to quantify because it is readily observable and simply defined. Aircraft handling is evaluated by assessing deviations from target aircraft parameters and configurations (e.g., Causse, Dehais, & Pastor, 2011; Transport Canada, 2017). For example, commercial pilots are expected to maintain predefined speeds and altitudes during certain phases of flight (Transport Canada, 2017). Deviations from target flight paths and aircraft configurations can lead to undesirable aircraft states, which increase the risk of incidents (Helmreich, 2000). Due to the reality of aviation operations, these deviations can be caused by the crew, or by external factors such as mission requirements, weather, or aircraft issues. Therefore, instead of focusing on avoiding deviations, flight crew are therefore trained to manage these deviations from ideal and reduce their impact on the flight. This decision-making relies on much more than simple manual flying skills. Each performance on a basic flying exercise builds on domain-specific knowledge and non-technical skills such as workload management and communication (Jensen, 1997).

Workload

Extensive research in human performance has investigated the relationship between operator states and performance. As pilot's primary role in flight operations is to perceive, process and act upon information from their environment, flight operations rely on the human operator's cognitive resources to accomplish this task (Endsley, 1995; Rolfe & Lindsay, 1973). A cognitive resource largely discussed in aviation is workload. Workload is a hypothetical construct that arises when a human operator must dedicate physical and psychological resources towards performance on a task (Hart & Staveland, 1988; Young et al., 2015). There is debate in the definitions of resources and task demands (see Cain, 2007), but most researchers acknowledge that workload arises when a human operator must accomplish a task. The interactions between an operator and their environment are part of a cycle where the human operator makes decisions that affect the task environment, and the task environment affects the human operator. This imposes costs on the human operator's resources, and these costs are defined as workload. In the present thesis, I have chosen to adopt Hart and Staveland's (1988) definition that workload reflects the "cost incurred by a human operator to achieve a particular level of performance". Workload reflects factors from the task, the operator and the interaction between them. Workload causes both a subjective experience within the operator, as well as physiological changes (Hart & Staveland, 1988; Lee & Liu, 2003).

A basic association between operator workload and operator performance is relatively well established. It has been demonstrated that operator performance is poor when workload is either too low or too high (Young et al., 2015). De Waard (1996) proposed that performance errors differ by their origin during periods of extreme low or extreme high workload. During extreme low, task demands are so low that operator's cognitive resources and attention can be diverted away from the task at hand. During extreme high workload, task demands are too high, and operators are unable to perceive, process or understand all relevant stimuli to perform adequately. Importantly, both extreme high and extreme low workload situations during flight training do not meet the intensity criteria for deliberate practice, limiting a trainee's learning possibilities (Ericsson, 2008).

Identifying the right intensity for training is difficult as workload, by nature, cannot be perceived by a training pilot in the context of flight training. Instructors therefore observe a learner's flight path and make inferences as to the psychological processes that led to the learner's performance. These inferences about underlying skills or processes can be inaccurate, jeopardising the effectiveness of feedback provided to the trainees. For example, Krusmark et al. (2004) found low interrater reliability between trained subject matter experts evaluating fighter pilots during a combat mission. Additionally, the subject matter experts' evaluations were unable to discriminate between different components of each mission. This jeopardizes the usefulness of any feedback provided by the instructors. Identifying differences between task components is essential for feedback to be accurate. Hence, increased understanding of a pilot's subjective perception and psychophysiological state during training would enable instructors to better understand a trainee's performance and identify directions for improvement. This would enable instructors to provide informed and immediate feedback prior to the next repetition of an exercise and target exercises which pilots would benefit most from repeating. More broadly, proper use of psychophysiological measures during flight training could ensure more time is spent conducting deliberate practice, as opposed to simple repetitions. Accomplishing this would require reliable measures of subjective experience and physiological state. For example, including psychophysiological measures of workload during flight training would enable instructors to identify points where a trainee's workload was high, thus suggesting the need for further training in that area. Increased attention to areas that require more improvement would enable trainees to conduct more deliberate practice during training, as opposed to simple repetition of tasks already within their capabilities.

De Waard proposed a series of properties that a measure must have in order to be feasible for an environment like aviation (de Waard, 1996 for details). Flight training requires indicators of workload in order to make within-person observations in a naturalistic environment. Therefore, four criteria are of particular interest: operator acceptance, sensitivity, primary task intrusion, and implementation requirements. First, pilots undergoing training should accept the measurement technique as being valid and beneficial. Second, sensitivity describes the property of a measure to detect changes in workload. Third, primary task intrusion should be kept to a minimum, meaning that the measure should not affect a pilot's primary task of flying the aircraft. Last, the technique must meet practical requirements. Equipment must be available, affordable, compatible with the cockpit environment, and operators must have the knowledge to interpret the information provided by the measures.

Generally, psychophysiological state in cognitive science has been quantified with subjective measures (e.g., NASA TLX), and physiological measures such as heart rate, eye tracking, electroencephalography (EEG) and secondary tasks (e.g., odd ball tasks). However, electroencephalography and secondary tasks do not meet the criteria of primary task intrusion and implementation requirements. Traditionally, EEG equipment is invasive and susceptible to electromagnetic interference from the multiple radio devices and electronics within aircraft cockpits. It also requires trained operators to install the electrodes in specific locations (e.g., using the 10-20 location system), which may be a time-consuming process and require expertise incompatible with commercial flight training. While it has been proposed that EEG would provide a useful measurement (e.g., Sterman & Mann, 1995), it has not been adopted as a main method for studying cognitive state in aviation. As portable EEG devices are becoming less intrusive and have better ambient electrical shielding, this may change.

Secondary task performance is a workload estimation technique where a subject is asked to perform a secondary task in addition to the primary tasks, such as responding to a tone by pressing a button. Performance on the secondary task (i.e., time to respond to the tone) is used to estimate the workload imposed by the primary task (Casner & Gore, 2010; de Waard, 1996). By design, this technique interferes with the primary task, often leading to dual-task effects that decrease performance in the primary task. Therefore, the present thesis focuses on the application of subjective workload, pupillometry and heart rate in an advanced flight training environment.

Subjective Workload

When an operator is accomplishing a task, the induced cognitive workload can be perceived by the operator. Humans can subjectively evaluate the workload imposed by a task. They are also capable of discussing their experience and assigning numerical values to those states (Gopher & Braune, 1984). Therefore, workload can be discussed or assessed through questionnaires (i.e., Hart & Staveland, 1988). These are recognized for being inexpensive, simple and having high face validity (Cain, 2007), thus meeting the criteria of implementation requirements and acceptance. However, subjective workload questionnaires are intrusive to the primary task. In aviation, the NASA Task Load Index (NASA-TLX) accomplishes this by collecting perceptions on six dimensions of workload: mental demand, physical demand, temporal demand, frustration, performance, and effort (Hart & Staveland, 1988). There is extensive literature supporting the implementation of NASA-TLX as a measure of workload in pilots (see Hart, 2006 for a review) . To be used in a training context, a measure of subjective workload should demonstrate a correlation with observed performance, and be sensitive to changes when exercises are repeated during a training session. In a study of fighter pilots conducting instrument approaches, where pilots approach to land without visual reference to the ground, within-subject differences in NASA TLX scores differentiated between best, medium and worst repetitions of the exercise (Mansikka et al., 2019a). Pilots subjectively reported low workload during their best performance. As their subjective workload increased, then their flight performance decreased, with reported workload at its highest during the pilots' worst approach (Mansikka et al., 2019a). These findings are similar in commercial airline pilots, where

NASA-TLX scores predicted performance evaluations from a pilot examiner (Lehrer et al., 2010). From an information processing perspective, these findings are expected, as automation of cognitive processes occur with learning and reduce workload on a given task (Newell et al., 1958). This is of particular importance to flight training, as it suggests that the NASA-TLX can be used to identify improvements in skills as during training. Nonetheless, self-report questionnaires like the NASA-TLX remains time-consuming (approx. 2 mins to complete) and invasive to the primary task. Therefore, research efforts have also investigated whether physiological measures can offer similar insight into workload without affecting the primary task.

Physiological Measures

By definition, workload occurs when a human operator engages and interacts with their environment. This interaction imposes costs on the human operator's resources, many of which occur concurrently with changes within the operator's physiology. Physiological sensors measure these changes, these data are then used to infer operator workload. Proponents of this approach highlight that they are not intrusive to the primary task, allowing individuals to perform in a naturalistic environment (Cain, 2007; Lehrer et al., 2010). For example, heart rate monitors are worn by athletes to monitor and optimize their training through an online measure of cardiovascular function. Recent developments in heart rate monitors have enabled manufacturers to produce accurate, affordable sensors. For example, many consumer grade smart watches include accelerometers, plethysmographs and ECG (e.g., Fitbit, Polar H10, Garmin Vivosmart, Apple Watch).

Cardiac function. There is significant literature demonstrating that cardiac function changes according to physical and psychological demands (reviewed in Roscoe, 1992). Many consumer heart rate monitors (e.g., Apple Watch 4) use an electrocardiograph (ECG). ECG measures electrical activity associated to the contractions of the heart muscle. An example of the latter is the Polar H10 (Polar, Kempele, Finland), which measures electrical activity using a chest strap. A normal ECG signal, showing voltage across time points, includes a sinus rhythm From there, two common measures can be extracted: heart rate and heart rate variability (HRV). Using the R peaks as a reference, heart rate is the number of R peaks per minute. HRV refers to the variability in the time interval between R peaks. These have both been investigated as indicators of workload (Cain, 2007).

This thesis adopts a holistic definition of workload, where workload's physiological responses are affected by both physical and mental demands. It is common knowledge that heart rate increases with physical activity. Heart rate also changes with many psychological states such as acute stress and fear (Åhs et al., 2009). There is also evidence that heart rate increases with cognitive load (Jorna, 1992; Luque-Casado et al., 2016). Therefore, heart rate could be considered as an indicator of workload. For example, heart rate was found to be higher in phases of flight associated with higher workload in pilots flying a single engine airplane (Causse et al., 2012). While there were no statistically significant differences between phases of flight, possibly due to the low sample size (n = 3), the data suggests practically relevant differences between each phase of flight. Heart rate was highest during take-off and landing, followed by climb and cruise, and lowest during cruise flight. This corresponds with task requirements from the various phases of flight. In another experiment, heart rate was collected for instrument approaches in pilots training for a qualification on the Lockheed P-3 Orion, a military aircraft similar to an airliner (Jorna, 1993). Heart rate increased progressively during the approach, corresponding to the increased task demands caused by increased precision required by pilots as they approach a runway. Additionally, heart rate during the approach was noticeably lower in pilots once they had completed the course, suggesting that the increased expertise lowered the workload experienced by pilots. Similar findings were observed in fighter pilots (Mansikka et al., 2015, 2016).

Another measure extracted from the ECG signal is heart rate variability (HRV), the variation in the duration of the R-R interval. As an indicator of the nervous system's control of cardiac function, it has been studied extensively in health research. Transient changes in HRV have also been linked to physiological processes, cognition and emotion (Thayer, 2009). Jorna (1992) proposed an early review of its validity as a measure of workload across domains, and HRV has since become a widely investigated measure (e.g., Cao et al., 2019; Charles & Nixon, 2019; Durantin et al., 2014; Jorna, 1993; Luque-Casado et al., 2016). It arises from the autonomic nervous system's control over the heart through its two branches: the sympathetic nervous system and the parasympathetic nervous system. The sympathetic nervous system increases heart rate, and is commonly associated with "fight, flight and or freeze" responses. The parasympathetic nervous system decreases heart rate, and is commonly associated with "rest and digest" (Thayer, 2009). Both systems are active concurrently (Quintana, 2019), and HRV is the result of their effect on heart rate.

Analyses of HRV are grouped into two domains: time-domain methods and frequency-domain methods (Shaffer & Ginsberg, 2017). HRV measures, like heart rate, are known to be affected by multiple physical and psychological processes, such as breathing (Jorna, 1992; Shaffer & Ginsberg, 2017). More importantly, there is substantial evidence suggesting that HRV changes with physical activity, attention, emotion and cognitive processing, all components of workload (Backs et al., 1994; Luque-Casado et al., 2016; Porges, 2007; Sandercock & Brodie, 2006; Thayer & Lane, 2000). In laboratory settings, there is evidence supporting that HRV changes in accordance to multiple physiological and psychological processes. Notably, successful emotion regulation has been linked to increased HRV (Butler et al., 2006). In both laboratory tasks and naturalistic environments, HRV has been linked with mental workload (Charles & Nixon, 2019). Of particular interest to training in safety-critical environments, police candidates who underwent situational awareness training both displayed increased performance and lower decreases in heart rate variability during simulated performance scenarios (Saus et al., 2006).

Considerable research has investigated using heart rate variability in the cockpit. De Rivecourt, Kuperus, Post, & Mulder (2008) conducted a study of pilot candidates in an instrument flying simulator. They asked participants to fly turns, altitude changes, and combinations during simulated instrument flight. Decreases in HRV from baseline and decreases in high frequency HRV were observed during manoeuvres with higher subjective and objective workload. However, Hankins & Wilson (1998) found few significant differences in HRV between flight segments, but did not collect data on pilot performance. In a study where performance was measured, interesting trends were found in fighter pilots conducting a series of manoeuvres in a simulator-based instrument proficiency check. Comparing pairs of manoeuvres, Mansikka et al. (2015) found that all pairwise comparisons with significant performance differences had significant differences in HRV. Required and perceived performance is an important component of workload, as it affects HRV through emotional processes and cognitive load. Taken together, these studies suggest that HRV may be a relevant indicator of pilot workload. However, additional research is required to clarify the relationship between these changes, pilot workload, and flying performance from an applied perspective. Specifically, emotion regulation, cognitive processing and attention are all involved in aircraft operations; but, are within-person changes in HRV during flight training related to operator performance? One of the aims of the current thesis is to address this question.

Pupillometry. Pupillometry has been considered as a measure of workload in cognitive science (see Laeng & Alnaes, 2019 for a review). Pupil size is generally measured with the use of an eye tracker, which can be fixed or portable. Fixed eye trackers are installed on a surface ahead of a pilot, facing the pilot. Portable eye trackers are generally mounted on frames similar to eyeglasses. In both cases, these unobtrusive devices use eye-facing cameras which detect and measure the pupil's size and orientation. In fact, both fixed and portable eye trackers have been successfully integrated in modern aircraft cockpits for engineering design and training purposes (e.g., Thales HuMans).

It is established that pupil size varies inversely to brightness through the pupillary light reflex (e.g., Laeng & Alnaes, 2019). Pupil diameter can also be affected by input from the autonomous nervous system. This supports its use as an indicator of workload when illumination is controlled for. Similar to HRV, increases in pupil size are associated with sympathetic activation, while decreases are associated with parasympathetic activation. Specifically, pupil diameter increases with cognitive and emotional processing (Van Der Wel & Van Steenbergen, 2018). Kahneman & Beatty (1966) found that pupil diameter increases with increased working memory load using a number-recall task. This relationship also seems to correlate with performance. Peavler (1974) demonstrated that individual differences in pupil diameter corresponded to differences in recall performance. In a complex task, Van Orden, Limbert, Makeig, & Jung (2001) collected pupillometry data during simulated anti-air-warfare control. Participants had to differentiate friendly from enemy aircraft, and make decisions as to whether to let aircraft pass or attack the aircraft. Workload was manipulated by increasing target density, thus requiring participants to process more information, and make more decisions. Error rates increased linearly as targets increased. Pupil dimeter followed the same linear trend. The pattern of these results provides evidence for an association between pupil diameter and workload.

Pupillometry has been used in several environments similar to flight operations. The difficulty in naturalistic environments lies with changes in illumination affecting pupil size though the pupillary light reflex (Dehais et al., 2010). The Index of Cognitive Activity (ICA) has been developed in order to compensate for this effect (Marshall, 2002). The ICA attempts to differentiate the rapid pupil changes caused by cognitive activity from the slower, more consistent changes caused by the pupillary light reflex. There is evidence supporting use of the ICA in environments similar to flight operations. In air traffic controllers, ISA and raw pupil diameter were related to objective workload, defined as a number of required tasks to accomplish (Rodríguez et al., 2015). In the same study, pupillometry was also related to subjective workload, except when workload was perceived as low. Schwalm, Keinath, & Zimmer (2008) asked drivers to drive a car simulator, making lane changes every 10 seconds under various secondary-task conditions. ICA scores were sensitive to the different conditions, but were also able to identify the momentary increases in workload during the lane changes. Furthermore, changes in ICA between conditions mirrored changes in performance. Similar studies in driving simulators observed the same relationship between workload and pupil size (ICA) (see Dlugosch, Conti, & Bengler, 2013; Palinko, Kun, Shyrokov, & Heeman, 2010). Taken together, these studies suggest that pupillometry can be a useful measure of workload.

Current Thesis

Modern flight operations require pilots to be trained as experts on the aircraft they operate, expertise which is built through effortful deliberate practice. Providing instructors with insight into pilots’ workload could enable them to better tailor their training to candidates’ current state, thus maximizing deliberate practice. Research in aviation and similar domains suggest that heart rate, heart rate variability, and pupil size can serve as indicators of workload. However, no studies have collected cardiac function, pupillometry, subjective workload and performance data in commercial pilots, limiting applicability of existing knowledge in training contexts. Therefore, the present thesis examines the relationship between physiological indicators of workload, subjective measures of workload, and basic flying performance.

The present thesis investigates the following hypotheses:

1. Heart rate, heart rate variability, and pupillometry changes with workload changes from manoeuvres of different difficulty.
2. Heart rate, heart rate variability and pupillometry can predict pilots' subjective workload ratings during aircraft handling exercises.
3. Workload can predict aircraft handling performance in commercial pilots.

Method

Participants

Participants (n = 14, males n = 12, Age Range 23-42, Median = 26.5) were recruited from the Montreal aviation community through flight operations managers and online pilot groups. To participate, pilots were required to hold a commercial (n = 5) or an airline pilot license (n = 9), with a valid medical certificate, and a current Transport Canada instrument flying rating. Pilots were currently employed in air taxi and aerial work (n = 5), commuter (n = 2) or airline transport (n = 7) operations. 9 were graduates of an aviation college or university program. Most participants were first officers (n = 8), with the remainder being captains.

Materials

Concordia Vision Lab Aviation Questionnaire

Participants were asked to fill out the Concordia Vision Lab Aviation Questionnaire. The questionnaire was designed to collect data about the pilots that will be used in future analysis of individual differences of results. Data collected includes the participants' flying qualifications, experience, education, and current pilot employment (for full version of the questionnaire, see Appendix 1). As this was the first use of the questionnaire, and due to the limited sample size, the questionnaire does not have any psychometric properties.

Flight simulator

All flying portions of this study were conducted in an Ascent XJ flight training device (Mechtronix, Montreal, Canada), which has been certified by Transport Canada for pilot training and evaluations. The Ascent XJ is a fixed base simulator with 180 degrees of visual display and a full size, two-crew cockpit. For this study, it was configured to represent a transport jet similar to the Boeing 737. Avionics included glass cockpit, dual flight management systems, and an autopilot controlled through a mode control panel. The simulated aircraft did not have an auto-throttle. Pilots were not expected to be familiar with specifics of the Ascent XJ's avionics, so all onboard equipment was pre-set for the experiment's scenario. The aircraft was configured to a standard weight of 60,000kg, and the "fuel consumption freeze" option was selected to ensure aircraft characteristics remained stable throughout the study.

Pilot monitoring

By design, commercial jets are designed to be operated by two pilots: a pilot flying and a pilot monitoring. In the experiments outlined below, an experimenter acted as pilot monitoring. An experimenter acting as pilot monitoring held commercial pilot licenses with qualifications for the class of airplane being operated. The participant pilot flying is responsible for conducting all flying tasks, notably operation of all flight controls. The pilot monitoring is a scripted role, where they also performed physical actions at the explicit request of the pilot flying and make certain “standard calls”. An example of a physical action included lowering landing gear on approach to landing following the “gear down” request. Standard calls are pre-defined verbalizations of information at certain points during flight, such as “80 knots” once the aircraft reaches that speed during take-off. Additionally, the role of the pilot monitoring was to advise the pilot flying of any parameter deviation past a limit set out in the Transport Canada Flight Test Guide (Transport Canada, 2017). Should there be a deviation, the pilot monitoring simply called out the parameter which exceeded the limit. For example, should a pilot be too high, the pilot monitoring would call “altitude”. As pilots had slight differences in phraseology and technique due to their employer’s operating procedures, the experimenter performed the pilot monitoring duties in accordance with the participant’s standard operating procedures.

Psychophysiological equipment

All psychophysiological data were recorded onto a Google OnePlus 6 A6003 running Android 10 (Google, Mountain View, USA). Eye tracking data were collected with a monocular Pupil Core head-mounted eye tracker (Pupil Labs, Berlin, Germany, and the custom Pupil Mobile application. The Pupil Core glasses were connected to the OnePlus via USB-C cable. The Pupil Core records both the right eye movements (192 x 192 pixel camera at a 200Hz sampling rate), and records the participant’s world view using a camera mounted above the eye (1080 x 730 pixels and a sampling rate of 30Hz).

Electrocardiogram data were collected using a Polar H10 Bluetooth heart rate monitor and the Elite HRV app (Polar Electro, Kempele, Finland; Elite HRV LLC, Asheville, USA) connected to the OnePlus via Bluetooth. The Polar H10 is a chest strap electrocardiogram with a sampling rate of 1000Hz. The Elite HRV app export consisted of sequential R-R interval durations in milliseconds. Temporal alignment was via an Android smartphone's internal clock.

Subjective workload

Subjective workload was measured through the NASA Task Load Index (NASA-TLX: Hart & Staveland, 1988). The NASA-TLX is a multidimensional workload scale, where participants answer six questions on a line ranging from 0 to 100. The questions refer to the various dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. Participant's responses on each dimension and the overall score were recorded via the NASA-TLX application on an Apple iPad (NASA, Mountain View, USA; Apple, Cupertino, USA). While the NASA-TLX does allow participants to complete a series of pairwise comparisons to weigh each dimension, the present study did not include this component.

Procedure

In the 24h period preceding the simulator session, participants completed the consent form and flight questionnaire at home. They also viewed a video briefing explaining study objectives, procedures, and simulator handling characteristics. Upon participants' arrival, participants were familiarized with the heart rate monitor, eye tracker and flight simulator. They were also introduced to an experimenter who was to be their pilot monitoring. Particular attention was paid to ensuring participants were comfortable with all flight controls and systems of the flight simulator, and the manoeuvres to be executed, as they would in preparing a real flight. The participants then put on the heart rate monitor and eye tracker, and took a seat in the simulator. All participants flew in the captain's seat. A second experimenter monitored and operated the flight simulator an instructor station outside of the simulator. All simulated flights were conducted at Montreal Trudeau International Airport (CYUL) on Runway 06 Left, using the required navigation procedures (NAV CANADA, Ottawa, Canada).

Familiarization Flight

The objective of this flight is to familiarize participants with the simulator's handling characteristics, and ensure harmony of procedures between the participant and pilot monitoring. Between each maneuver, the autopilot is engaged while the participant delivers a briefing on how they will conduct the upcoming maneuver. Once ready, the participant disengages the autopilot and begins the maneuver.

The familiarization flight begins at the runway threshold, where participants complete their take off briefing, before the take-off checklist is completed. Participants then conduct a visual take off and establish cruise flight at 5000 feet, as directed by the experimenter. Once participants declare they are comfortable with the handling of the aircraft, they proceeding to a series of standardized maneuvers including a medium turn, then an approach to stall in clean configuration. Once these maneuver exercises were completed, the simulator was repositioned for an instrument landing system (ILS) approach. The familiarization flight ended once the aircraft was stopped on the runway. Participants were offered a 10-minute break after that flight, during which any remaining questions were answered.

Experiment Flight

Prior to the experiment flight, the heart rate monitor and eye tracker were turned on and calibrated. To allow for a baseline resting state comparison, we collected a 3-minute ECG baseline. In addition, we calibrated the eye tracker with Pupil Labs' portable calibration target. The experiment flight was similar to the practice flight, except for three additions. The air exercises included a steep turn, and an approach to stall in landing configuration. These are similar exercises to the medium turn and approach to stall in the familiarization flight, but with higher difficulty. Pilots are familiar with both these exercises, as are part of Transport Canada's Pilot Proficiency Check Flight Test (Transport Canada, 2017). Second, participants conducted the ILS (instrument landing system) approach, where pilots precisely follow vertical and horizontal guidance to the ground, twice at the end of the flight. One was uneventful, while the other approach included a laser strike prior to crossing the final approach fix. The approaches were counterbalanced, and the simulator was repositioned between approaches. Between each maneuver, the airplane was put on autopilot while participants completed a NASA-TLX subjective workload questionnaire. Participants themselves disengaged the autopilot when ready for the next maneuver. The study ended once the airplane was immobilized. Participants returned the psychophysiological equipment and were provided an opportunity for questions. Durations for each manoeuvre and flight varied based on pilots' speeds. Overall recording times for the manoeuvres was 995 seconds (*SD* = 159). Normal turn duration was (*M* = 225, *SD* = 24) seconds, steep turn duration was (*M* = 158, *SD* = 36) seconds, clean approach to stall duration was (*M* = 111, *SD* = 7) seconds, and landing configuration approach to stall duration was (*M* = 110, *SD* = 30) seconds.

Data Handling

Due to the small participant sample size, missing or unreliable data was excluded as interpolation or mean substitution would be unreliable. Heart rate data was missing for one participant. When there was presence of >20% unreliable or missing data for pupillometry data, pupillometry was excluded on a per-manoeuvre basis. NASA-TLX subjective workload ratings were missing for two participants.

Electrocardiogram

Interbeat interval data was exported from the EliteHRV application into Kubios HRV 3.3.1 (Kubios, Kuopio, Finland) for cleaning and frequency analysis. Artifact correction was conducted through the built-in algorithm, using a medium threshold. Visual inspection was then conducted for quality assurance. Analysis windows were created around each maneuver, from the time the autopilot was disengaged to the point it was turned on after completion or the aircraft was stopped after landing. Heart rate and the standard deviation of beat-to-beat intervals (SDNN) were calculated for each manoeuvre.

Eye Tracking

Eye tracking data handling was done in Pupil Player 1.17.6 (Pupil Labs, Berlin, Germany). As we used the mobile recorder, post-hoc pupil detection, calibration and gaze mapping were conducted with the built-in tools using standard settings. The data was then exported for processing through a custom R script. There, samples with a confidence lower than .6, the standard Pupil Lab cut-off, and extreme pupil diameters (+/- 3 SD from the mean) were marked as missing. This was done because many of these stretches correspond to blinks. Then, missing values in stretches shorter than 500ms, where lost pupil diameters can be expected to remain relatively constant, were interpolated with a linear interpolation function. Stretches longer than 500ms were left as missing. From there, means and standard deviations were calculated for each manoeuvre. Manoeuvres where more than 80% of the data was missing were excluded from further analysis.

Simulator Data

Unlike in a traditional lab environment, we cannot access the simulator data as a direct output due to Transport Canada certification. Consequently, the simulator data was extracted via video recording of a screen displaying simulator parameters. Values from these videos were extracted every second via optical character recognition using a custom R program implementing Google's Tesseract OCR engine (Google, Mountain View, USA). Results from the Tesseract OCR interface were manually checked for accuracy by comparing them to the video frames. Incorrect values were replaced manually. Time windows were then created around each maneuver: from the time the autopilot was disengaged to the point it was turned on after completion or the airplane was stopped after landing. For each time window (i.e. manoeuvre), airspeed and altitude data was extracted for analysis.

Results

Data analyses were conducted through JASP (Version 0.13) and R (Version 3.6.3). Due to a pandemic caused by the SARS-CoV-2 coronavirus, in-person research was halted as a preventative measure on March 17th, 2020 in response to government public health measures. Therefore, we were unable to attain the planned sample size (> 20 participants) as recommended by the power analysis. The present analyses were conducted on the existing sample at the time of lab closure (n = 14). When conducting statistical tests on small samples that compare an alternate hypothesis to a null hypothesis, the lack of statistical power is an issue that can cause decision errors (Kline, 2013). This means that traditional parametric statistical tests are less likely to detect the presence of a significant effect, even if one is present. Additionally, use of statistical tests with excessively small samples increases the chance of falsely rejecting the null hypothesis when it should have been maintained (Kline, 2013). Therefore, the data analysis below will consist of two sections. Firstly, descriptive statistics and qualitative descriptions of data distributions will be discussed. Second, exploratory statistical analyses will be presented, with an emphasis on Bayes Factors, effect sizes, and confidence intervals. These three types of analysis are less prone to small sample size / low power issues in comparison to traditional parametric analysis (Kline, 2013).

Descriptive Statistics

Descriptive statistics and distributions plots were calculated for flight path deviations, heart rate, heart rate variability, pupil size and NASA-TLX ratings (see Table 1 & Figure 1). Qualitative observation of these plots highlights similarities within pilots. Within-subject patterns are observable in heart rate, SDNN, and mean pupil diameter (see figure 1). These statistics and observations are discussed in this section.

Flight path deviations

For turns, flight path deviations refer to the total time, in seconds, spent outside acceptable speed (240 knots – 260 knots) and altitude (4900 feet – 5100 feet) (Transport Canada, 2017). For stalls, flight path deviations refer to the time, in seconds, for participants to regain safe airspeed (200 knots) and recover to an acceptable altitude (4800 feet), as time

Table 1

Descriptive Statistics for heart rate, heart rate variability (SDNN), mean pupil diameter, flight path deviations and NASA-TLX

	Heart Rate				SDNN				Mean Pupil Diameter				Flight Path Deviations				NASA-TLX			
	CAS	NT	SAS	ST	CAS	NT	SAS	ST	CAS	NT	SAS	ST	CAS	NT	SAS	ST	CAS	NT	SAS	ST
n	13	13	13	13	13	13	13	13	8	9	10	10	14	14	14	14	12	12	12	12
M	81.21	82.70	79.20	84.47	40.15	30.62	37.96	29.98	5.51	5.26	5.65	5.49	43.50	6.50	17.50	13.29	50.49	40.56	37.99	49.58
SD	7.03	7.52	7.34	7.63	6.30	10.59	9.59	9.04	1.17	0.83	1.20	0.76	11.95	12.60	5.37	18.47	16.26	14.69	13.33	16.39
Min.	65.18	65.12	63.86	64.06	29.63	17.47	25.88	15.89	3.38	4.27	3.14	4.61	32.00	0.00	12.00	0.00	9.17	16.67	15.83	11.67
Max.	89.35	94.85	89.70	93.37	52.01	44.35	59.27	42.28	6.92	6.67	6.91	6.80	67.00	39.00	33.00	68.00	66.67	67.50	55.00	70.83

Note. CAS: complex approach to stall, NT: normal turn, SAS: simple approach to stall, ST: steep turn

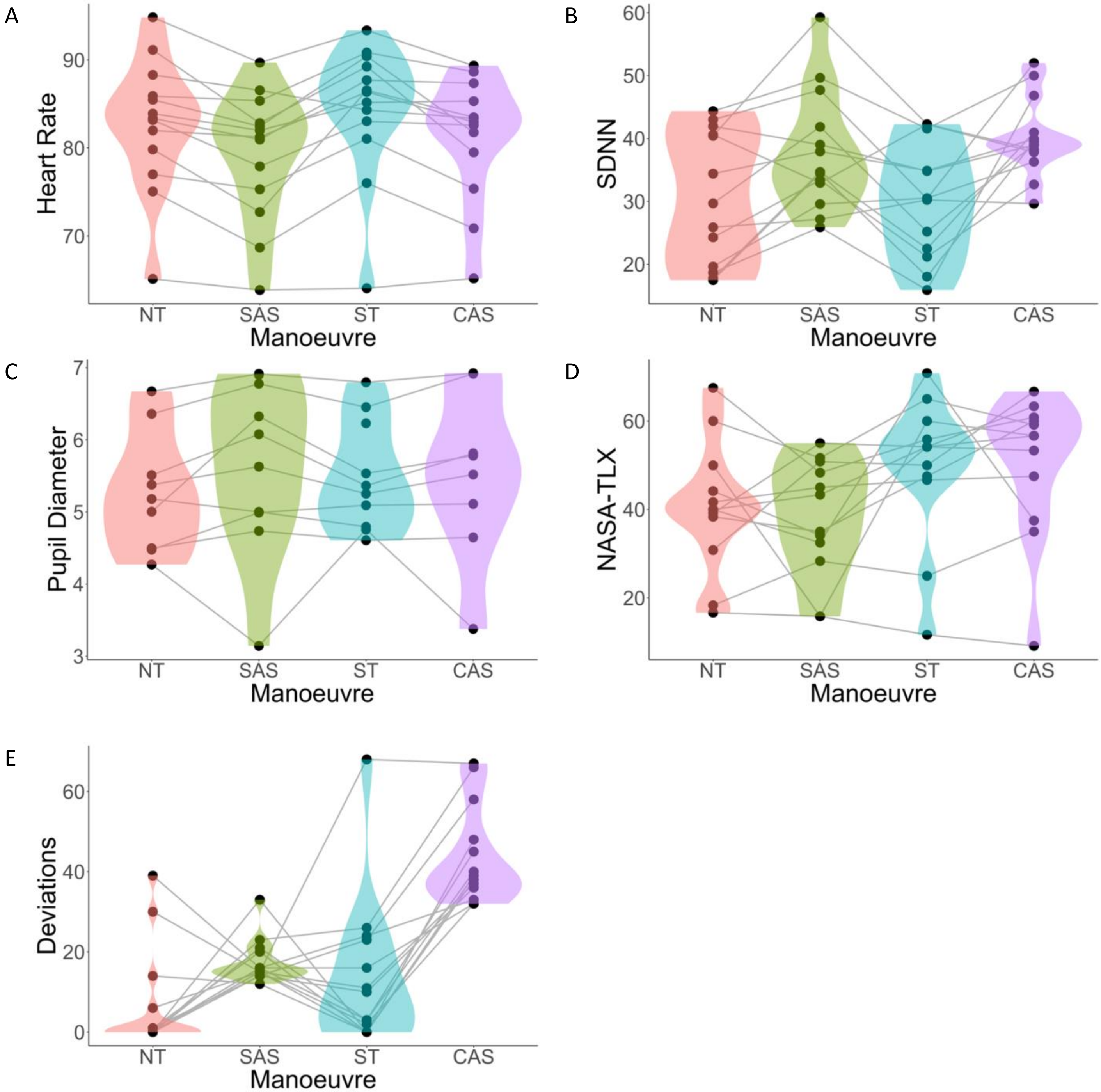

Figure 1 Distribution plots for each measure, by manoeuvre

(A) Heart rate in beats per minute **(B)** Heart rate variability (SDNN) in milliseconds **(C)** Mean pupil diameter, in arbitrary Pupil Labs units **(D)** Compound NASA-TLX subjective workload ratings, in percentage **(E)** Time outside ideal flight path, in seconds.

Each dot represents a participant, and lines connect within-subject observations.

Note: NT: normal turn, SAS: simple approach to stall, ST: steep turn, CAS: complex approach to stall outside these tolerances is associated to increased danger. Sample sizes, mean and standard deviations were calculated for flight path deviations for each manoeuvre. Lower flight path deviations indicate superior performance.

Mean flight path deviations were higher during steep turns (*M* = 13.29, *SD* = 18.47) than during normal turns (*M* = 6.50, *SD* = 12.60). Observations of the descriptive plot (Figure 1) shows deviations on both manoeuvres to be positively skewed with (*n* = 8) pilots remaining within acceptable tolerances during the entirety of the normal turn, and (*n* = 4) pilots remaining withing acceptable tolerances during steep turn.

Mean recovery time for the simple (i.e., clean configuration) approach to stall was (*M* = 17.50, *SD* = 5.37). Mean recovery time for the complex (i.e., landing configuration) approach to stall was (*M* = 43.50, *SD* = 11.95). Pilots required between 12 and 33 seconds to recover from the simple approach to stall, while for complex approach to stall required 32 to 67 seconds to recover from the stall. Qualitative observations of the distribution plots (Figure 1) for both approach to stall show that there is a larger variance in performance on the complex approach to stall (SD=11.95) compared to the simple approach (SD=5.37).

Objective workload

Objective workload analyses focused on heart rate (in beats-per-minute), heart rate variability (i.e. SDNN) and pupil size (in arbitrary units). Means and standard deviations were calculated for each manoeuvre. One participant was missing from heart rate analyses due to an equipment malfunction, therefore heart rate and SDNN analyses were conducted on a sample size of (*n* = 13). Participants were excluded from the pupillometry analyses on a per-manoeuvre basis; thus, sample size will be reported for each manoeuvre and measure.

Mean heart rate during the steep turn was 84.47 (*SD* = 7.36), while during the normal turn it was 82.70 (*SD* = 7.34). Mean heart rate during the complex approach to stall was 81.21, (*SD* = 7.03), and slightly lower during the clean stall (*M* = 79.20, *SD* = 7.52). Qualitative observation of the distribution plots (see Figure 1) suggest similar variability in the distributions between both difficulties of turns. During approach to stall, the distributions appear slightly negatively skewed and slightly lower than heart rate during the turns. However, heart rate distributions for both difficulties remain similar.

Mean SDNN during was similar between steep turns was 29.98 (*SD* = 9.04) and normal turns (*M* = 30.62, *SD* = 10.59). During approaches to stalls, SDNN was slightly higher than during turns (complex: *M* = 40.15, *SD* = 6.30; simple *M* = 37.96, *SD* = 9.69). On the harder approach to stall, variance was lower than during other manoeuvres (*SD* = 6.30). Observations of the distribution plots highlight this statistic, with the distribution for the complex approach to stall noticeably different from the other manoeuvres (see figure 1).

For each manoeuvre, mean pupil size were calculated (see table 1). Group means, standard deviations, and sample sizes were calculated for manoeuvre mean and median pupil sizes (see table 1). Qualitative observation of the distribution plots highlights different distributions between approaches to stall and turns (figure 1). There is lower variance in mean pupil size during approaches to stall than during turns. Distributions for turns are positively skewed, contrary to distributions for approaches to stall, which are negatively skewed. Meanwhile, no differences are perceivable between manoeuvre difficulty (i.e. comparing both turns and stalls).

Subjective workload

Descriptive statistics for subjective workload scores were calculated for each factor of the NASA-TLX (table 2). Each factor of the NASA-TLX measures a dimension of workload (i.e. mental demand, physical demand, temporal demand, effort, frustration and perceived performance). Total NASA-TLX scores, with all factors equally weighted, were also calculated as a multidimensional estimate of workload (Hart & Staveland, 1988), see table 1. Distribution plots suggest higher subjective workload for the difficult turn and approach to stall when compared to the easier manoeuvres, but ratings (i.e. NASA-TLX) were much less consistent between maneuvers than were objective measures. Additionally, subjective workload on the difficult manoeuvres appears to be negatively skewed, suggesting a larger proportion of responses indicating higher perceived workload (see figure 1).

Table 2

Means and standard deviations of NASA-TLX components by manoeuvre.

	Mental		Physical		Temporal		Performance		Effort		Frustration	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
NT	53.3	18.7	52.1	22.7	27.5	24.9	35.0	18.3	50.4	16.2	25.0	19.2
SAS	45.0	20.7	41.3	24.5	45.8	25.7	31.3	14.3	43.3	18.4	21.3	14.3
ST	58.3	18.5	55.0	18.6	37.9	24.6	47.5	21.9	60.4	17.9	38.3	27.5
CAS	59.6	18.8	50.0	21.4	46.7	25.3	47.5	20.4	59.6	18.8	39.6	21.6

Note: NT: normal turn, SAS: simple approach to stall, ST: steep turn, CAS: complex approach to stall

Exploratory analyses

Due to the limited sample size, the following statistical tests are considered exploratory and should therefore be interpreted with caution. In small samples, p-values used to make decisions on hypotheses are unreliable, thus increasing the probability of decision errors (Kline, 2013). This affects all statistical tests relying on p-values, including tests designed to check assumptions (i.e. Shapiro-Wilk test of normality, Mauchly's test of sphericity). Therefore, we assume that assumptions regarding hypothesis testing are violated. Therefore, the present exploratory analyses focus on effect sizes, Bayes Factors and confidence intervals. P-values will not be reported. Although statistical significance can be inferred from confidence intervals, interpretation of statistical significance is avoided in the context of these analyses. Interpretations of Bayes Factors follow general guidelines, which are limited in applicability until a review of evidence in human factors research provides further guidance in this field (Wetzels et al., 2011).

Between groups

Five exploratory repeated-measures analyses of variance (ANOVA) were conducted to investigate between-manoeuvre differences in performance, objective measures of workload, and NASA-TLX subjective workload. Planned post-hoc, uncorrected, dependent samples *t*-tests are reported to compare exercise pairs, contrasting the difficult (i.e. steep turn and complex approach to stall) and easier (i.e. normal turn and simple approach to stall) versions of each manoeuvre. When observation of the ANOVA plots suggested a difference between exercise pairs, post hoc *t*-tests were conducted to further investigate these differences and calculate an effect size.

First, an ANOVA was conducted to investigate overall performance differences between manoeuvres, with manoeuvre (turn, stall) as a within-subject factor, which suggests a large effect of manoeuvre on performance, $F(3,39) = 29.37$, $\eta^2 = .69$, with decisive evidence for the research hypothesis ($BF_M = 2.44e+8$). Two non-directional, dependent sample *t*-tests were conducted to compare means between steep turn ($M = 13.29$, $SD = 18.47$) and normal turns ($M = 6.50$, $SD = 12.60$), and between complex approach to stall ($M = 43.50$, $SD = 11.95$) and simple approach to stall ($M = 17.50$, $SD = 5.38$). For turns, anecdotal evidence was found suggesting no difference in performance between normal and steep turns ($BF_{10} = .73$), with a medium effect size (*d* = -.42, 95% CI [-.96, .14]). For approaches to stall, decisive evidence was found suggesting a difference in performance between the complex and simple manoeuvres (BF_{10} = 10511.67), with a large effect size (*d* = -2.19, 95% CI [-3.15, -1.19]).

Second, an ANOVA was conducted to investigate overall heart rate differences between manoeuvres, which suggests an effect of manoeuvre on heart rate, $F(3,36) = 9.97$, $\eta^2 = .45$, with decisive evidence for the research hypothesis (BF_M = 324.58). Two non-directional, dependent sample *t*-tests were conducted to compare means between steep turn (*M* = 84.47, *SD* = 7.63) and normal turns (*M* = 82.70, *SD* = 7.52), and between complex approach to stall (*M* = 81.21, *SD* = 7.03) and simple approach to stall (*M* = 79.20, *SD* = 7.52). For turns, anecdotal evidence was found suggesting an anecdotal increase in heart rate between normal and steep turns (BF_{10} = 1.15), with a medium effect size (*d* = .53, 95% CI [-.06, 1.11]). For approaches to stall, anecdotal evidence was found suggesting an anecdotal increase in heart rate between the simple and complex manoeuvres (BF_{10} = 1.09), with a medium effect size (*d* = .52, 95% CI [-.07., 1.09]).

Third, an ANOVA was conducted to investigate overall differences in SDNN between manoeuvres, which suggests a moderate effect of manoeuvre on SDNN ($F(3,36) = 8.70$, $\eta^2 = .42$) with decisive evidence for the research hypothesis (BF_M = 164.82). Two non-directional, dependent sample *t*-tests were conducted to compare means between steep turn (*M* = 29.98, *SD* = 9.04) and normal turns (*M* = 30.62, *SD* = 10.59), and between complex approach to stall (*M* = 40.15, *SD* = 6.30) and simple approach to stall (*M* = 37.96, *SD* = 9.59). For turns, anecdotal evidence was found suggesting no difference in SDNN between normal and steep turns (BF_{10} = .30), with a negligible effect size (*d* = -.11, 95% CI [-.44, .65]). For approaches to stall, anecdotal evidence was found suggesting no difference in SDNN between the complex and simple manoeuvres (BF_{10} = .38), with a small effect size (*d* = -.44, 95% CI [-.79., .32]). However, qualitative observation of the descriptives plot suggests differences between approach to stall exercises and turn exercises (see figure 1). Dependent, non-directional *t*-tests were therefore conducted contrasting the normal turn and simple approach to stall, and the steep turn with complex approach to stall. Substantial evidence was found suggesting that the SDNN may be lower during the normal turn than the simple approach to stall (BF_{10} = 9.22) with a large effect size (*d* = -.93, 95% CI [-1.58, .26]). The mean for the complex manoeuvres suggested substantial evidence that SDNN is lower during the steep turn than the approach to stall (BF_{10} = 9.91), with a large effect size (*d* = -.95, 95% CI [-1.59, -.27]).

Fourth, an ANOVA was conducted to investigate overall differences in mean pupil diameter between manoeuvres, which suggests a no effect of manoeuvre on mean pupil diameter (F(3,21) = .93, η^2 = .12) with substantial evidence for the null hypothesis (BF_M = 3.06). These results and qualitative observation of the plot did not warrant further pairwise comparisons as no between-manoeuvre differences and patterns were observable (see figure 1).

Last, an ANOVA was conducted to investigate overall differences in subjective workload between manoeuvres, which found substantial evidence against differences in subjective workload across manoeuvres (F(3,33) = 5.24, η^2 = .32, BF_M = 9.78). Two non-directional, dependent sample *t*-tests were conducted to compare means between steep turn (*M* = 48.82, *SD* = 16.07) and normal turns (*M* = 41.88, *SD* = 15.26), and between complex approach to stall (*M* = 50.49, *SD* = 16.26) and simple approach to stall (*M* = 37.99, *SD* = 13.33). For turns, substantial evidence was found suggesting a difference in subjective workload between normal and steep turns (BF_{10} = 4.22), with a large effect size (*d* = .83, 95% CI [.15, 1.48]). For approaches to stall, substantial evidence was found suggesting a difference in subjective workload between the complex and simple manoeuvres (BF_{10} = 9.24), with a moderate effect size (*d* = .99, 95% CI [.27, 1.67]).

Predicting subjective workload and performance

Exploratory linear regressions were used to predict subjective workload and performance from the objective workload measures.

First, heart rate, SDNN and mean pupil diameter were regressed onto NASA-TLX subjective workload ratings across all manoeuvres combined. Anecdotal evidence was found against the overall model (F(3,25) = 1.37, BF_{10}= .35), which had a small effect size (R^2 = .14). Confidence intervals for all coefficients in this model included zero. Model summary and coefficients are available in table 3.

Table 3

Model summary and coefficients for predicting NASA-TLX subjective workload ratings from heart rate, SDNN and pupil diameter.

	B	*95% CI* for *B*		$BF_{Inclusion}$	*F*	*df*	BF_{10}	R^2
		LL	*UL*					
Overall Model					1.37	3, 25	.35	.14
Intercept	73.68	-64.38	211.75	1.00				
Heart Rate	-.51	-1.67	.64	.80				
SDNN	.43	-.73	.82	.47				
Pupil Diameter	2.17	-4.97	9.32	.53				

Next, regressions were conducted to predict flight path deviations from workload measures. First, heart rate, SDNN and mean pupil diameter were regressed onto flight path deviations across all manoeuvres. Anecdotal evidence was found for the overall model, regressing heart rate, SDNN and mean pupil diameter onto performance ($F(3,33) = 3.03$, BF_{10} = 1.48). This model had a small effect size (R^2 = .22). Second, NASA-TLX ratings were regressed onto flight path deviations. Strong support was found for the overall model ($F(1,46) = 9.77$, BF_{10} = 12.44), which had a small effect size (R^2 = .18). Last, subjective and objective workload measures were used to predict flight path deviations. Strong evidence was found for the overall model ($F(4,24) = 5.37$, BF_{10} =15.30), which had a medium effect size (R^2 = .47). In the compound model, Bayes Factors for all coefficients suggested some evidence for their inclusion in the model (all $BF_{Inclusion}$'s > 1). Model summaries and coefficients are in table 4.

Table 4

Model summaries and coefficients for predicting flight path deviations from psychophysiological measures (heart rate, SDNN and pupil diameter), subjective reports (NASA-TLX) and a combined model with all measures.

	B	*95% CI*		$BF_{Inclusion}$	*F*	*df*	BF_{10}	R^2
		LL	*UL*					
Psychophysiological					3.03	3,33	1.48	.22
Intercept	-37.38	-141.98	67.22	1.00				
Heart Rate	.51	-.48	1.50	.79				
SDNN	.96	.26	1.65	3.18				
Pupil Diameter	-3.18	-9.39	3.02	.80				
Subjective Workload					9.77	1,46	12.44	.18
Intercept	-1.78	-17.22	13.65	1.00				
NASA-TLX	.51	.18	.83	12.44				
Composite Model					5.37	4,24	15.30	.47
Intercept	-157.54	-309.08	-6.00	1.00				
Heart Rate	1.45	.20	2.72	5.12				
SDNN	1.11	.28	1.94	6.64				
Pupil Diameter	-1.50	-9.22	6.22	1.46				
NASA-TLX	.72	.28	1.16	12.95				

Discussion

The present thesis aimed to investigate the usability of psychophysiological measures, specifically heart rate, heart rate variability, and pupillometry in flight training contexts in commercial pilots. To do this, three specific hypotheses were put forward. First, we hypothesized that heart rate, heart rate variability and pupil size will vary as a function of the change in workload caused by different flight manoeuvres. Second, we hypothesized that heart rate, heart rate variability and pupillometry can be used to predict a pilots' subjective workload ratings during aircraft handling exercises. Last, we hypothesized that subjective workload can predict aircraft handling performance. Here I will review the evidence regarding these hypotheses, discusses limitations for the present study, and provides recommendations for future directions.

Between-Manoeuvre Workload Differences

A valid psychophysiological measure of workload should be able to differentiate between high and low workload manoeuvres. This was our first hypothesis, for which the data suggest partial support. We found differences in the cardiac measures, NASA-TLX ratings and flight path deviation patterns when comparing manoeuvre types. However, we did not observe a difference in pupillometry across manoeuvre type.

To check that manoeuvre manipulations were effective at creating a low and high workload, we confirmed that the complex variant of each manoeuvre (i.e., turn and approach to stall) was more difficult than the simple manoeuvres, indicated by increased deviations from the ideal flight path. We found that flight path deviations were higher in the complex manoeuvres than the deviations during simple manoeuvres. Transport Canada (2004) classifies steep turns and landing configuration approach to stalls as complex variants of the simple manoeuvres, requiring increased skill and resources from pilots for acceptable completion of the manoeuvre. Our results provide evidence that support this statement, as lower performance was observed during the complex manoeuvres, and variance in performance across pilots was higher during the complex manoeuvres. Presence of workload differences between manoeuvres were also confirmed by examination of subjective workload as assessed via the NASA-TLX scores. Steep turns and complex approaches to stall were rated as more demanding than the simple variants. Taken together, the observed performance and subjective workload differences between manoeuvres suggests an effective manipulation of pilots' workload between manoeuvres. However, do the changes in workload caused by the different manoeuvres cause a corresponding change in the cardiac function and pupillometry?

Based on findings from Jorna (1993) and Luque-Casado (2016), increases in heart rate and decrease in heart rate variability (measured by SDNN) is associated with increased arousal and workload. Therefore, we expected to observe in the complex exercises an increase in heart rate, and SDNN to decrease. However, our observations differed from this prediction. We observed clear differences between manoeuvre types (i.e., turn vs stall), but little difference was found between the complex and simple variant of each manoeuvre. Specifically, pilots' heart rate was higher during the turn manoeuvre than during the approaches to stall. Heart rate variability was lower during turns. When contrasting the complex approach to stall to its simple variant, and the steep turn to the normal turn, the data shows small increases in heart rate with the increase in difficulty. Similar findings were observed with heart rate variability. Heart rate variability was highest during stalls than during turns. Given the limited sample size, we were unable to conduct traditional parametric statistical tests on these between-condition differences. Nevertheless, the general pattern of the results for cardiac measures partially supports our hypothesis by differentiating between manoeuvre types (i.e. turns vs. approaches to stall), but not between different difficulties of the same manoeuvre.

While we were unable to conduct statistical testing due to the small sample size, these results do replicate the findings reported in similar studies. For example, a study conducted in light aircraft found increases in heart rate during high workload conditions, such as take-off and landing (Causse et al., 2012). As with our results, the increase in heart rate during more demanding manoeuvres were stable between pilots, despite having large between-subject variance. Hankins and Wilson (1998) conducted a similar experiment that collected psychophysiological data during an entire flight while pilots performed multiple manoeuvre types (i.e. turns, cruise flight, holding patterns). They found differences in heart rate (in beats per minute) across manoeuvres types, but they did not observe with changes in heart rate variability. Taken together with the results of the present experiment, heart rate appears to change predictably due to workload induced by different manoeuvre types, but not vary as a function of workload changes due to manoeuvre difficulty. The one caveat to this conclusion is that the manipulation of difficulty may not have been sufficient to create a significant change in heart rate (see workload and performance limitations later).

Moving on to the pupillometry data, we aimed to investigate whether changes in pupil diameter due to workload were detectable in a flight simulator environment. In naturalistic environments, changes in brightness remains a confound to measuring workload through pupil size. However, in a simulated environment, brightness in the environment is expected to be consistent between the simple and complex manoeuvres. We therefore hypothesized that mean pupil size during a manoeuvre would change between manoeuvres and differences in workload. Our data did not support this hypothesis. No differences were present in pupil size between manoeuvre types, nor between levels of difficulty. These results contrast those reported by Dehais and colleagues (2010), who found differences in pupil size between phases of flight during simulated emergencies at night. This is likely due to the fact that the cockpit environment was subject to variations in brightness across different areas of the cockpit that render pupillometry data without a precise ambient light measurement. Dehais and colleagues (2010) tightly controlled brightness by conducting their study at night and only considering samples with equal ambient light. Given our results in a controlled, simulated environment, we interpret our results to conclude that pupil size as a measure of workload remains inappropriate for the cockpit environment, as proper measurement would require experimental control beyond those practical for flight training. While overall brightness did not change significantly during a simulator session, our pupillometry measures may be affected further by factors such as technique and attention. For example, some pilots conducting these manoeuvres may have directed their attention towards areas in their environment of varying brightness, spending more time looking at elements inside the dark cockpit than time spend looking outside the forward windows. However, various transformations of the pupillometry signal remain untested in the cockpit environment. These are discussed further in this chapter.

Taken together, the findings reported here suggest that additional research is required to define contexts where psychophysiological measures are most appropriate measurements of workload. For example, the changes in cardiac function related to workload imposed by basic processes required to complete a task, regardless of its difficulty level. In other words, one could hypothesize that the patterns of between-manoeuvre changes in heart rate and heart rate variability are caused by the inherent differences in pilot resources required for completing each manoeuvre. Turns require sustained attention, consistently updating a model in working memory with multiple parameters (i.e. bank angle, altitude, airspeed, engine thrust) and precision. Conversely, stalls require the prompt application of a trained rehearsed recovery procedure, with an emphasis placed on quickly and successfully exiting the situation, and not on the precision of flying inputs. Changes in psychophysiological measures may be due to these different types of task demands. Specifically, heart rate would be lower during stalls because these require a simpler, highly rehearsed recovery instead of an extended increase in attention and precise movements of the flight controls during turns. Meanwhile, subjective workload evaluations are often biased by previous exposure to similar manoeuvre and performance (Moore & Picou, 2018). In this case, participants evaluating a given manoeuvre may be evaluating this manoeuvre to previous exposure to similar situations, regardless of the basic resources required for completing a given manoeuvre type. This may explain subjective workload scores' ability to differentiate between manoeuvre difficulty levels, as flight path deviations were generally higher during the difficult manoeuvre, but not manoeuvre type. The present results add to the literature supporting the strategic use of cardiac function as an objective workload measure to compare aviation situations likely to require different types of cognitive and physiological resources on behalf of the pilot.

Psychophysiological measures, subjective workload and performance

The second aim of this thesis was to investigate the association between physiological measures as an objective measure of workload, between workload and aircraft handling performance. Due to the limited statistical power of the current study, the preliminary conclusions regarding these hypotheses are speculative. Two hypotheses were put forward to investigate the association between psychophysiological measures, and between workload and flying performance.

First, we hypothesized that heart rate, heart rate variability and pupil diameter could predict subjective workload. In general, our data did not support this hypothesis. Differences between manoeuvres were present in subjective workload and psychophysiological measures, but did not correlate. Contrarily to the changes in heart rate between manoeuvre types, we found that pilots subjectively evaluated the difficult manoeuvres as more demanding than the simple ones. Meanwhile, subjective workload did not vary by manoeuvre type, as did heart rate. As with the results for our first hypothesis, these findings suggest that subjective workload and psychophysiological measures are distinct indications of the mental and physiological resources required by a pilot in order to complete an aircraft handling task.

Moving towards relating workload and performance, we tested whether workload could predict aircraft handling performance. At first glance, our results did not support this hypothesis. We find that heart rate, heart rate variability and pupil diameter accounted for a small proportion of the variance in performance (22%). Similarly, subjective workload ratings accounted for a small proportion of the variance in performance (18%). However, when taken together, objective and subjective measures of workload accounted for a moderate proportion of the variance in flight path deviations (47%). Based on the results from our second hypothesis, the present experiment provides evidence supporting the use of psychophysiological measures in conjunction to subjective perceptions of workload, in order to better understand pilot workload. This dissociation between objective and subjective workload had been highlighted by Luque-Casado et al. (2016) during laboratory tasks. They reported that NASA-TLX subjective workload ratings suggested a different pattern of changes in workload than suggested by heart rate variability during laboratory cognitive tasks (i.e.: n-back task, duration discrimination task, psychomotor vigilance task). But a clear understanding of the link between subjective workload, objective workload and performance does not yet exist.

By definition, workload is intertwined with performance, as performance requirements will affect the amount of cognitive and physical resources required by an operator to complete a task (Endsley, 1995; Hart & Staveland, 1988). Measuring workload generally relies on subjective ratings or psychophysiological measures, each with their own caveats. Subjective workload ratings are biased by perceived performance in a task (Moore & Picou, 2018). Meanwhile, psychophysiological measures often measure associated constructs such as arousal, which are believed to be related to performance by an inverse-U curve. This theory is often a derivative of the Yerkes-Dodson law (see Cain, 2007; Causse et al., 2012). Causse et al.

(2012) proposed that due to the nature of aviation operations, where significant performance degradations are highly unlikely due to extensive flight training and emphasis on safety, most aviation research captures only the increasing, first half of the U-curve. Given the small sample sizes in aviation research, the association between workload and performance in aviation research can easily appear to be a linear, but imperfect, association. However, no studies to date have provided conclusive evidence for this proposition. Nevertheless, evidence for the inclusion of workload measures in flight training to provide feedback continues to accumulate. For example, a series of studies with Finnish fighter pilots has suggested the inclusion of cardiac function as an objective workload measurement supplementing traditional feedback, where heart rate variability changed across manoeuvre types and between repeated performances of a manoeuvre during advanced flight training (Mansikka et al., 2015; 2016; 2019). The present thesis adds to this body of evidence, suggesting a link between objective workload, subjective workload measures and performance. Furthermore, the present study also suggests that a combination of objective and subjective workload measures would increase potential usefulness in an aviation training context until further research clarifies the associations between each measure and performance.

Limitations and recommendations

First, commercial pilots are a relatively small population, forcing us to minimize data loss and exclusions. This was further exacerbated by the COVID-19 pandemic, which unexpectedly prevented the completion of data collection in March of 2020. Thus, the observed sample size was smaller than planned, collecting data with 14 pilots instead our anticipated sample size of 25 pilots. To compensate for this, data cleaning in this thesis was aimed at minimizing data exclusions. Specifically, exclusions were conducted on a per-manoeuvre basis, and only the measures with excessive missing data were excluded. This means that many manoeuvres were represented in two measures (i.e., cardiac measures and subjective workload), while the others (i.e., pupillometry) was excluded for this manoeuvre. Pupillometry data was particularly excluded by this decision, where thresholds for missing data were set considerably lower than guidelines for research (Kline, 2013). Nonetheless, our final sample size was not uncommon for most published aviation research. Generally, research in aviation using psychophysiological measures is conducted with sample sizes between 10 and 30 (e.g., Hankins & Wilson, 1998;

Hidalgo-Muñoz et al., 2018; Lee & Liu, 2003; Mansikka et al., 2019). However, results with sample sizes of (n = 3) have been published (Causse et al., 2012), as even small sample sizes provide guidance for further research.

A consequence of the small sample sizes and objectives of this research area is that statistical power is difficult to achieve. As such, assumptions regarding traditional parametric null-hypothesis statistical tests are often violated, and cannot be effectively tested (Kline, 2013). In psychophysics, this is exacerbated by the large variation in psychophysiological measures due to measurement error and extraneous physiological mechanisms. Furthermore, different transformations of psychophysiological signals often measure different phenomena. For example, different heart rate variability metrics, while all transformations of an electrocardiography signal, may produce differing results (Cain, 2007; Roscoe, 1992). Our exploratory statistical analyses and associated implications are therefore limited. To compensate, the current research focused on effect sizes, Bayes Factors, descriptive statistics and qualitative observations of plots. An increased sample size would enable effective statistical analysis (Kline, 2013). In the meantime, considering the "case study" nature of how workload measure in flight training are used (e.g., Thales Humans), future research could focus on reviewing existing studies to identify consistent patterns in psychophysiological responses between exercises and studies. An increased understanding of these consistent patterns would then have to be related to changes in performance, or learning, to provide tangible benefits.

Psychophysiological measures. A principle objective of the present experiment was to maximise ecological validity. Thus, integration of psychophysiological measures into a cockpit environment implies certain challenges, especially in a research environment. The cardiac measures and pupillometry used in this study collect data about physiological processes in a participant, but causality between those physiological changes and psychophysiological measures are always assumed. Confounding factors during the data collection can alter these psychophysiological results. For example, heart rate variability is known to follow the circadian rhythm, and pupillometry is known to be affected by ambient light – both extraneous factors not related to pilot workload. Researchers and practitioners must refrain from inferences about causal relationships and relying on multiple measures (i.e. observed performance and subjective measures) in order to better interpret psychophysiological measures. Our pupillometry findings serve as a good example of this, where we did not observe meaningful changes in pupil size. Given the naturalistic environment, this was expected. However, future pupillometry research could investigate derivatives of the pupillometry measures such as the Index of Cognitive Activity (Marshall, 2002; Schwalm et al., 2008). Future research, enabled by portable eye tracking, could also investigate the use of pupil diameter in concordance with gaze mapping to control for brightness of the cockpit area where pilots are looking. With respect to cardiac measures, the changes we observed must also be interpreted with caution. Use of cardiac function as an indicator of pilot workload does not allow for diagnostic of workload: where the source for the workload change is identified. For example, we cannot infer whether it was mental workload, stress, or physical workload that caused the observed increases heart rate during turns, when compared to stalls.

Our results were based on the comparisons of physiological measures between different manoeuvres, due to the lack of an acceptable baseline. In psychophysiological research, contrasting measures during a experimental manipulations to those from baseline is common practice (Butler et al., 2006; Shaffer & Ginsberg, 2017). This aims to control for individual characteristics not due to experimental manipulations. Some studies elect to contrast cardiac function during a manipulation to an individual's resting state, forming a work-rest ratio (Causse et al., 2011). However, one's level of activation is known to changed based on environment, so baselines should be obtained within a representative environment, such as straight and level cruising flight (e.g., de Waard, 1996). However, our baseline was taken in the cockpit, with the airplane on the runway with engines running and parking brake on. The objective was to emulate the condition found in the cockpit, but without any task-related workload. Initial investigation of these data highlighted large differences across participants baselines. Some participants' baseline heart rate was higher than all other manoeuvres, suggesting that participants felt that measurement period as "pre-takeoff", and they were already conducting mental task-related planning. We therefore elected to analyze the current data by only comparing psychophysiological data across manoeuvres. This imposes limitations on the comparisons we can make. In the future, the inclusion of multiple baseline measurements could provide several advantages. Accumulation of multiple baselines would enable recommendations to be made regarding appropriate conditions for baseline. Also, multiple baselines would allow results from a study to be better contrasted to the existing literature, where baseline measurements are not identical across studies.

Workload and Performance. By definition, workload during a task varies according to the required performance level. Thus, pilots may have felt that performance in this study was not important, as there were no consequences of substandard performance. However, this limitation is expected that this translate to training situations, where one's performance expectations may moderate the subjective and physiological response during a manoeuvre. Yet, the relationship between workload and performance in aviation remains to be clarified.

It has been proposed that most aviation research only investigates a limited range of the inverse-U shaped association between workload and performance, where performance increases with workload (Causse et al., 2012). From an information processing perspective, however, one could expect workload in a given task to decrease with expertise. This happens as expertise decreases the resources required to complete a task through more effective mental models, which automate behaviours and decision-making. This makes sense, as aviation research has focuses on normal and trained emergency procedures with highly trained personnel, it is therefore difficult to create ecologically valid scenarios to challenge trained experts. Little research is devoted to investigating workload and performance at, and beyond, the upper limits of a pilot's competence (Orlandi & Brooks, 2018; Wang et al., 2016; Young et al., 2015). Yet, the breakdown of human performance during high workload is of major interest for human factors scientists. In fact, it has been dubbed the human "redline" by researchers (Young et al., 2015). In the current study, we observed variance in performance to be much higher in the complex stall than other manoeuvres, where flight path deviations were kept to a minimum. It is possible that our most of our manoeuvres did not induce enough workload to increase heart rate significantly, limiting our observations to only a small segment of possible workload levels. In order to better understand the association between workload and performance, future research efforts could be directed towards identifying patterns in objective and subjective workload indicators in situations beyond pilots' current competence. For example, a continuation of this current project could involve unusual attitudes: where an airplane is unexpectedly forced into an aggressive position requiring immediate, and prompt, recovery from the pilot to ensure a positive outcome. In sum, investigating the limits of human performance in aviation though a combination of subjective and objective measures will help scientists understand the measurement of workload, and its relationship to performance in complex, naturalistic environments.

Conclusion

We aimed to understand the association between subjective measures of workload, heart rate, heart rate variability, pupillometry and pilot performance. The present study contributes to the human factors literature by applying subjective and objective measures of workload in commercial pilots conducting aircraft handling exercises. Despite the unexpectedly low sample size due to a global pandemic, three main findings arose from the present experiment. First, heart rate and heart rate variability patterns were stable between participants, showing clear differences between manoeuvre types, but not manoeuvre difficulty. Second, pupil diameter is not an effective measure of workload in the simulator cockpit, likely due to the differences in brightness. Third, psychophysiological measures of workload can supplement subjective workload reports to better understand performance. Despite this, many questions remain unanswered before indicators of workload can be effectively implemented during training scenarios. Of particular interest are the causal relationships in these measures, especially during periods of extremely high workload where performance begins to decrease. That being said, human factors scientists are eager to better understand the role of workload in training and operations safety. Future research and replication are required to clarify the associations between psychophysiological measures, subjective workload and flying performance.

References

Åhs, F., Sollers Iii, J. J., Furmark, T., Fredrikson, M., & Thayer, J. F. (2009). High-frequency heart rate variability and cortico-striatal activity in men and women with social phobia. *NeuroImage*, *47*, 815–820. https://doi.org/10.1016/j.neuroimage.2009.05.091

Backs, R. W., Ryan, A. M., & Wilson, G. F. (1994). Psychophysiological Measures of Workload during Continuous Manual Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *36*(3), 514–531. https://doi.org/10.1177/001872089403600308

Bazargan, M., & Guzhva, V. S. (2011). Impact of gender, age and experience of pilots on general aviation accidents. *Accident Analysis and Prevention*, *43*, 962–970. https://doi.org/10.1016/j.aap.2010.11.023

Butler, E. A., Wilhelm, F. H., & Gross, J. J. (2006). Respiratory sinus arrhythmia, emotion, and emotion regulation during social interaction. *Psychophysiology*, *43*(6), 612–622. https://doi.org/10.1111/j.1469-8986.2006.00467.x

Cain, B. (2007). *A Review of the Mental Workload Literature*. https://apps.dtic.mil/dtic/tr/fulltext/u2/a474193.pdf

Cao, X., MacNaughton, P., Cadet, L. R., Cedeno-Laurent, J., Flanigan, S., Vallarino, J., Donnelly-McLay, D., Christiani, D. C., Spengler, J., Allen, J. G., Cao, X., MacNaughton, P., Cadet, L. R., Cedeno-Laurent, J. G., Flanigan, S., Vallarino, J., Donnelly-McLay, D., Christiani, D. C., Spengler, J. D., & Allen, J. G. (2019). Heart Rate Variability and Performance of Commercial Airline Pilots during Flight Simulations. *International Journal of Environmental Research and Public Health*, *16*(2), 237. https://doi.org/10.3390/ijerph16020237

Casner, S. M., & Gore, B. F. (2010). *Measuring and Evaluating Workload: A Primer*. https://www.researchgate.net/profile/Brian_Gore/publication/260284948_Measuring_and_Evaluating_Workload_A_Primer/links/00b7d5307a650e725f000000.pdf

Causse, M., Dehais, F., Faaland, P.-O., & Cauchard, F. (2012). An analysis of mental workload and psychological stress in pilots during actual flight using heart rate and subjective measurements. *Proceedings of the 5th International Conference on Research in Air Transportation (ICRAT '12)*.

Causse, M., Dehais, F., & Pastor, J. (2011). Executive Functions and Pilot Characteristics Predict Flight Simulator Performance in General Aviation Pilots. *THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY*, *21*(3), 217–234. https://doi.org/10.1080/10508414.2011.582441

Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, *74*, 221–232. https://doi.org/10.1016/J.APERGO.2018.08.028

De Rivecourt, M., Kuperus, M. N., Post, W. J., & Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, *51*(9), 1295–1319. https://doi.org/10.1080/00140130802120267

de Waard, D. (1996). *The Measurement of Drivers ' Mental Workload* (Issue 4) [University of Groningen]. https://doi.org/10.1016/j.apergo.2003.11.009

Dehais, F., Causse, M., & Pastor, J. (2010). *Toward the definition of a pilot's physiological state vector through oculometry: a preliminary study in real flight conditions*. https://www.semanticscholar.org/paper/Toward-the-definition-of-a-pilot's-physiological-a-Dehais-Causse/ddecb1105a75fee9df2a6d80ec51ead9eed87a79

Dlugosch, C., Conti, A. S., & Bengler, K. (2013). Driver Distraction through Conversation Measured with Pupillometry. *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 198–204. https://doi.org/10.17077/drivingassessment.1488

Duckworth, A. L., Kirby, T. A., Tsukayama, E., Berstein, H., & Ericsson, K. A. (2011). Deliberate Practice Spells Success: Why Grittier Competitors Triumph at the National Spelling Bee. *Social Psychological and Personality Science*, *2*(2), 174–181. https://doi.org/10.1177/1948550610385872

Durantin, G., Gagnon, J., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural Brain Research*, *259*, 16–23. https://doi.org/10.1016/J.BBR.2013.10.042

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2006). Expertise and Situation Awareness. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 633–651). Cambridge University Press.

Endsley, M. R. (2000). Situation Models: An Avenue to the Modeling of Mental Models. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, "Ergonomics for the New Millennium,"* 61–64. https://doi.org/10.1177/154193120004400117

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, *15*(11), 988–994. https://doi.org/10.1111/j.1553-2712.2008.00227.x

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*(8), 725–747. https://doi.org/10.1037/0003-066x.49.8.725

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/doi:10.1037/0033-295X.100.3.363

Gopher, D., & Braune, R. (1984). On the Psychophysics of Workload: Why Bother with Subjective Measures? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*(5), 519–532. https://doi.org/10.1177/001872088402600504

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEC and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, *69*(4), 360–367.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Issue 52, pp. 139–183). Elsevier Science Publishers.

Helmreich, R. L. (2000). On error management: lessons from aviation. *BMJ : British Medical Journal*, *320*(7237), 781. https://doi.org/10.1136/BMJ.320.7237.781

Hidalgo-Muñoz, A. R., Mouratille, D., Matton, N., Causse, M., Rouillard, Y., & El-Yagoubi, R. (2018). Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation. *International Journal of Psychophysiology*, *128*, 62–69. https://doi.org/10.1016/J.IJPSYCHO.2018.04.002

ICAO. (2017, November 29). *ICAO NGAP Summit addresses pressing shortages of skilled professionals for future air transport network*. *November*. https://www.icao.int/Newsroom/Pages/ICAO-NGAP-Summit-addresses-pressing-shortages-of-skilled-professionals-for-future-air-transport-network.aspx

Jensen, R. S. (1997). The Boundaries of Aviation Psychology, Human Factors, Aeronautical Decision Making, Situation Awareness, and Crew Resource Management. *The International Journal of Aviation Psychology*, *7*(4), 259–267. https://doi.org/10.1207/s15327108ijap0704

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

Jorna, P. G. A. M. (1993). Heart rate and workload variations in actual and simulated flight. *Ergonomics*, *36*(9), 1043–1054. https://doi.org/10.1080/00140139308967976

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585. https://www.jstor.org/stable/1720478

Kellogg, R. T., & Whiteford, A. P. (2009). Educational Psychologist Training Advanced Writing Skills: The Case for Deliberate Practice. *Educational Psychologist*, *44*(4), 250–266. https://doi.org/10.1080/00461520903213600

Kennedy, Q., Taylor, J. L., Reade, G., & Yesavage, J. A. (2019). Age and Expertise Effects in Aviation Decision Making and Flight Control in a Flight Simulator. *Aviation, Space, and Environmental Medicine*, *39*(5), 41. https://doi.org/10.3357/ASEM.2684.2010

Kline, R. (2013). *Beyond Significance Testing* (2nd Editio). American Psychological Association. https://doi.org/10.1037/14136-000

Krusmark, M., Schreiber, B. T., & Bennett, W. (2004). *The Effectiveness of a Traditional*

Gradesheet for Measuring Air Combat Team Performance in Simulated Distributed Mission Operations. https://apps.dtic.mil/dtic/tr/fulltext/u2/a428119.pdf

Laeng, B., & Alnaes, D. (2019). Pupillometry. In C. Klein & U. Ettinger (Eds.), *Eye Movement Research An Introduction to its Scientific Foundations and Applications* (pp. 449–496). Springer Nature Switzerland. https://doi.org/https://doi.org/10.1007/978-3-030-20085-5

Lee, Y. H., & Liu, B. S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, Space, and Environmental Medicine*, *74*(10), 1078–1084.

Lehrer, P., Karavidas, M., Lu, S., Vaschillo, E., Vaschillo, B., & Cheng, A. (2010). Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory study. *International Journal of Psychophysiology*, *76*(2), 80–87. https://doi.org/10.1016/j.ijpsycho.2010.02.006

Luque-Casado, A., Perales, J. C., Cárdenas, D., & Sanabria, D. (2016). Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological Psychology*. https://doi.org/10.1016/j.biopsycho.2015.11.013

Mansikka, H., Simola, P., Virtanen, K., Harris, D., & Oksama, L. (2016). Fighter pilots' heart rate, heart rate variation and performance during instrument approaches. *Ergonomics*, *59*(10), 1344–1352. https://doi.org/10.1080/00140139.2015.1136699

Mansikka, H., Virtanen, K., & Harris, D. (2019a). Comparison of NASA-TLX scale, modified Cooper-Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics*, *62*(2), 246–254. https://doi.org/10.1080/00140139.2018.1471159

Mansikka, H., Virtanen, K., & Harris, D. (2019b). Comparison of NASA-TLX scale, modified Cooper-Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics*, *62*(246–254). https://doi.org/10.1080/00140139.2018.1471159

Mansikka, H., Virtanen, K., Harris, D., & Simola, P. (2015). Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test. *Applied Ergonomics*, *56*, 213–219. https://doi.org/10.1016/j.apergo.2016.04.006

Marshall, S. P. (2002). The Index of Cognitive Activity: Measuring Cognitive Workload. *IEEE 7th Human Factors Meeting*, 7.5-7.9. www.eyetracking.com]

McKinney, E. H. J., & Davis, K. J. (2003). Effects of Deliberate Practice on Crisis Decision Performance. *Human Factors*, *45*(3), 436–444. https://journals.sagepub.com/doi/pdf/10.1518/hfes.45.3.436.27251?casa_token=WHaZA-AMdx0AAAAA:L-sN8nCMC9cRzDlVJ7vXbq3IP8AtPka-ip2aZq9F61VlGrkTdR_wlCsD1OSqtJZK127gQMzAdHQXr0c

Moore, T. M., & Picou, E. M. (2018). A potential bias in subjective ratings of mental effort. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2405–2421. https://doi.org/10.1044/2018_JSLHR-H-17-0451

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*(3), 151–166. http://search.proquest.com.ezproxy.ub.gu.se/docview/614294792/fulltextPDF/6C05A3B1F5F1407DPQ/3?accountid=11162

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall. https://psycnet.apa.org/record/1973-10478-000

Orasanu, J., & Fischer, U. (1997). Finding Decisions in Natural Environments: The View from the Cockpit. In C. Zsambok & G. A. Klein (Eds.), *Naturalistic Decisoin Making*. Lawrence Erlbaum Associates. https://www.researchgate.net/profile/Judith_Orasanu/publication/290608120_Finding_Decisions_in_Natural_Environments_The_View_from_the_Cockpit/links/569ae0e608ae6169e55daeba/Finding-Decisions-in-Natural-Environments-The-View-from-the-Cockpit.pdf

Orlandi, L., & Brooks, B. (2018). Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. *Applied Ergonomics*, *69*, 74–92. https://doi.org/10.1016/j.apergo.2018.01.005

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *Eye Tracking Research and Applications Symposium (ETRA)*, 141–144. https://doi.org/10.1145/1743666.1743701

Peavler, W. S. (1974). Pupil Size, Information Overload, and Performance Differences.

Psychophysiology, *11*(5), 559–566. https://doi.org/10.1111/j.1469-8986.1974.tb01114.x

Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, *74*(2), 116–143. https://doi.org/10.1016/j.biopsycho.2006.06.009

Quintana, D. (2019). *Heart rate variability: physiology, methodology and experimental possibilities*. University of Oslo. https://www.youtube.com/watch?v=TIf01TRMUpw&feature=youtu.be

Rodríguez, S., Sánchez, L., López, P., & Cañas, J. J. (2015). Pupillometry to assess air traffic controller workload through the mental workload model. *Proceedings of ATACCS 2015 - 5th International Conference on Application and Theory of Automation in Command and Control Systems*, 95–104. https://doi.org/10.1145/2899361.2899371

Rolfe, J. M., & Lindsay, S. J. E. (1973). Flight deck environment and pilot workload: Biological measures of workload. In *Applied Ergonomics*. https://pdf.sciencedirectassets.com/271441/1-s2.0-S0003687000X01391/1-s2.0-0003687073902159/main.pdf?x-amz-security-token=AgoJb3JpZ2luX2VjECUaCXVzLWVhc3QtMSJHMEUCIE8HuUlP70id%2BofcCfxdRRcptOGEFsZxbs042ATeX70zAiEA1OOTw1EExc%2B%2BpX53gjbtSX3qgMHmKLor6gFrHdQ

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, *34*(2–3), 259–287. https://doi.org/10.1016/0301-0511(92)90018-P

Sandercock, G. R. H., & Brodie, D. A. (2006). The use of heart rate variability measures to assess autonomic control during exercise. *Scandinavian Journal of Medicine and Science in Sports*, *16*(5), 302–313. https://doi.org/10.1111/j.1600-0838.2006.00556.x

Saus, E.-R., Johnsen, B. H., Eid, J., Riisem, P. K., Andersen, R., & Thayer, J. F. (2006). The Effect of Brief Situational Awareness Training in a Police Shooting Simulator: An Experimental Study. *Military Psychology*, *18*(Suppl.), S3–S21. https://doi.org/10.1207/S15327876MP1803S_2

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors*, *50*(6), 864–878. https://doi.org/10.1518/001872008X374974

Schwalm, M., Keinath, A., & Zimmer, H. D. (2008). Pupillometry as a method for measuring mental workload within a simulated driving task. In D de Waard, B. Flemisch, B. Lorenz, L.

Oberheid, & K. A. Brookhuis (Eds.), *Human Factors for assistance and automation* (pp. 1–13). Shaker Publishing. https://www.researchgate.net/profile/Maximilian_Schwalm/publication/262357754_Pupillometry_as_a_method_for_measuring_mental_workload_within_a_simulated_driving_task/links/0c960537600688a792000000/Pupillometry-as-a-method-for-measuring-mental-workload-with

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*(September), 1–17. https://doi.org/10.3389/fpubh.2017.00258

Standing Committee on Transport Infrastructure and Communities. (2019). *Supporting Canada's Flight Schools: Report of the Standing Committee on Transport, Infrastructure and Communities* (Issue April). https://www.ourcommons.ca/Content/Committee/421/TRAN/Reports/RP10406018/tranrp29/tranrp29-e.pdf

Sterman, M. B., & Mann, C. A. (1995). Concepts and applications of EEG analysis in aviation performance evaluation. *Biological Psychology*, *40*(1–2), 115–130. https://doi.org/10.1016/0301-0511(95)05101-5

Thayer, J. F. (2009). Heart Rate Variability: A Neurovisceral Integration Model. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 1041–1047). Elsevier. https://doi.org/10.1016/B978-008045046-9.01991-4

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, *61*(3), 201–216. https://doi.org/10.1016/S0165-0327(00)00338-4

Todd, M. A., & Thomas, M. J. W. (2012). Flight hours and flight crew performance in commercial aviation. *Aviation, Space, and Environmental Medicine*, *83*(8), 776–782. https://doi.org/10.3357/ASEM.3271.2012

Transport Canada. (2004). *Flight Instructor Guide - Aeroplane*. Government of Canada. https://doi.org/10.1037/h0034711

Transport Canada. (2017). *Pilot Proficiency Check and Aircraft Type Rating Flight Test Guide (Aeroplanes)* (Issue First Edition (Revision 1)). Government of Canada.

https://www.tc.gc.ca/eng/civilaviation/publications/tp14727-menu-2709.htm

Van Benthem, K., & Herdman, C. M. (2013). Individual pilot factors predict simulated runway incursion outcomes. *International Symposium on Aviation Psychology*, 579–584. https://corescholar.libraries.wright.edu/isap_2013/18

Van Der Wel, P., & Van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*, 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T.-P. (2001). Eye Activity Correlates of Workload during a Visuospatial Memory Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(1), 111–121. https://doi.org/10.1518/001872001775992570

Wang, L., He, X., & Chen, Y. (2016). Quantitative relationship model between workload and time pressure under different flight operation tasks. *International Journal of Industrial Ergonomics*. https://doi.org/10.1016/j.ergon.2016.05.008

Waterson, P. (2011). World War II and other historical influences on the formation of the Ergonomics Research Society. *Ergonomics Research Society*, *54*, 1111–1129. https://doi.org/10.1080/00140139.2011.622796

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on Psychological Science*, *6*(3), 291–298. https://doi.org/10.1177/1745691611406923

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. https://doi.org/10.1080/00140139.2014.956151

Appendix 1

Consent Form and Concordia Vision Labs Aviation Questionnaire

CVL - Aviation Form

CONCORDIA UNIVERISTY

INFORMATION AND CONSENT FORM

Study Title: Physiological correlates of human performance in aviation
Researchers: Sam Clement-Coulson
Researcher's Contact Information: sam.coulson@concordia.ca
Faculty Supervisor: Dr. Aaron Johnson
Faculty Supervisor's Contact Information: aaron.johnson@concordia.ca
Source of funding for the study: N/A

You are being invited to participate in the research study mentioned above. This form provides information about what participating would mean. Please read it carefully before deciding if you want to participate or not If there is anything you do not understand, or if you want more information, please ask the researcher.

A. PURPOSE
The purpose of the research is to understand the links between pilot experience, skill, performance and their physiological correlates.

B. PROCEDURES
If you participate, you will be asked to fill out this questionnaire, then fly maneuvers in a flight simulator while wearing recording equipment (Polar heart rate monitor and Pupil Labs eye tracking glasses). You will be aske for you feedback on your experience after the study. In total, participating in this study will take about 90 minutes.

C. RISKS AND BENEFITS
You might face certain risks by participating in this research. These risks include: eye strain, fatigue and simulator sickness. You will be offered breaks to minimize this risk.

D. CONFIDENTIALITY
We will gather the following information as part of this research: accuracy, response time, heart rate, eye tracking and aviation experience. We will not allow anyone to access the information, except people directly involved in conducting the research. We will only use the information for the purposes of the research described in this form, and for publication of the study results. The information gathered will be coded. That means that the information will be identified by a code. The researcher will have a list that links the code to your name. We will protect the information by keeping it in a separate file from the rest of the data saved on a password protected computer. We will destroy the information five years after the end of the study.

F. CONDITIONS OF PARTICIPATION
You do not have to participate in this research. It is purely your decision. If you do participate, you can stop at

any time. You can also ask that the information you provided not be used, and your choice will be respected. If you decide that you don't want us to use your information, you must tell the researcher before leaving the experiment room. There are no negative consequences for not participating, stopping in the middle, or asking us not to use your information. You can also ask that the information you provided not be used, and your choice will be respected. If you decide that you don't want us to use your information, please inform the researcher before leaving the experiment room. Should you decide to withdraw your information after the completion of data collection, you will have four days after leaving the experiment room to do so.

G. PARTICIPANT'S DECLARATION

I have read and understood this form. I have had the chance to ask questions and any questions have been answered. I agree to participate in this research under the conditions described.

If you have questions about the scientific or scholarly aspects of this research, please contact the researcher. Their contact information is at the top of this text. You may also contact their faculty supervisor.

If you have concerns about ethical issues in this research, please contact the Manager, Research Ethics, Concordia University, 514.848.2424 ex. 7481 or oor.ethics@concordia.ca
* Required

1. Name: *

____

2. Participant ID *

____

3. I have read and understood this form. I have had the chance to ask questions and any questio have been answered. I agree to participate in this research under the conditions described. *

Mark only one oval.

( ) AGREE
( ) DISAGREE

Briefing

Briefing

 http://youtube.com/watch?v=AHwWMYnHz90

Demographics

4. Gender *

Mark only one oval.

- ( ) Female
- ( ) Male
- ( ) Other: ____________________

5. Date of Birth *

____

Example: January 7, 2019

Licenses and Experience

6. What kind of airplane license do you currently hold? *

Mark only one oval.

- None
- Student Pilot/Recreational Permit
- PPL - Private Pilot License
- CPL - Commercial Pilot License
- ATPL - Airline Transport Pilot License

7. What ratings (airplane) do you hold? *

Check all that apply.

- [ ] Multi-Engine
- [ ] Instrument Flying
- [ ] Instructor
- [ ] Seaplane

8. List your type ratings, separated by commas

9. How many years has it been since your first solo? *

10. Are you attending, or did you complete, an aviation college/university program? *

Mark only one oval.

- Yes
- No

Employment

11. Select the most appropriate pilot employment status: *

Mark only one oval.

- Unemployed / training
- Flight Instructor
- Private operator pilot (604)
- Aerial work pilot (702)
- Air taxi pilot (703)
- Commuter pilot (704)
- Airline pilot (705)

12. What position do you normally fly in? *

Mark only one oval.

- Single Pilot or Instructor
- Relief First Officer
- First Officer
- Relief Captain
- Captain

CQFA Simulator

13. Have you been in this simulator? *

Mark only one oval.

- Never had
- 1 year ago
- 2 years ago
- 3 years ago
- 4 years ago
- 5 years ago
- >5 years ago

14. How many hours do you have in this simulator? *

Thank you!