Setting a baseline for cognitive fatigue in student pilots

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Aviation

at Massey University, Palmerston North, New Zealand

Nicholas John Ashley

2013

Abstract

Despite fatigue being an important topic in many areas of aviation, little is known about its occurrence and effects amongst student pilots. The purpose of this study is to examine cognitive fatigue in the flight training environment with the goal of setting a baseline for fatigue accumulation over the course of a one hour training flight.

The study was divided into three sections. Firstly, information was gathered on the numbers and distribution of student pilots across New Zealand, and this resulted in a decision that research would proceed with students at a single large flight training organisation. Next, a search was undertaken for pre-existing tools that could be modified and refined to be made suitable for use in the flight training environment. A questionnaire and reaction time test were then created and successfully validated in a pilot study. Finally the main body of the study comprised using the two tools to test a non-probability sample of 21 student pilots, split between a main group and a control group. The data were then collated and analysed to determine the level of fatigue which accumulated, assess correlations between variables, and evaluate the significance of the results.

Results were overall satisfactory, with the questionnaire returning some of the most useful and significant data. Significant levels of fatigue were detected amongst participants, but it could not be exactly determined how this would affect performance. Several significant correlations were discovered between different variables, which served to both reinforce existing knowledge on the topic, and further confirm the validity and reliability of the tools. While the study was somewhat limited in its approach and scope, it is relatively ground-breaking, and creates the potential for further research in this area.

Acknowledgements

I would like to acknowledge the contribution of my supervisor Dr Jose Pérezgonzález, his knowledge and guidance has been invaluable to the process of completing this study.

Contents

List of tables

List of figures

List of appendices

Introduction

Despite fatigue being a prominent consideration in many aviation sectors, it has received comparatively little attention in general aviation [GA]. With research in the specific area of flight training being particularly lacking, this study takes an initial step to develop tools and procedures, attempt to analyse fatigue levels in student pilots, and set a baseline from which future study can be built on. This study attempts to set such a baseline through exploratory research involving student pilots training towards a fixed-wing aeroplane pilots licence.

Chapter one, the literature review, establishes the foundation of the study by examining the complex nature of fatigue from four broad aspects: definitions, causes, effects, and measurement, followed by a brief overview of the study. Over 100 years of research into many aspects of fatigue has failed to produce much in the way of a widely accepted definition, and establishing an appropriate definition is a necessary initial step before proceeding further with the study. The causes of fatigue are also important considerations; there are many different factors, some of which only apply to specific situations. With many varying effects, fatigue has the potential to seriously compromise human performance, especially with regard to effective information processing. This can create serious safety deficiencies, which is particularly relevant in aviation, where major accidents continue to happen as a result of errors induced by fatigue. The second chapter covers a small sub-study that was undertaken to evaluate student pilots in New Zealand which served as a basis to determine the future scope and direction of the research. Next, chapter three briefly outlines information that was gathered on methods that have been used to measure fatigue in medical and aerospace environments. Drawing from this information, a questionnaire was designed and validated alongside a reaction time test obtained from a psychology software package. Chapter four details the main section of the research, involving testing methodology, data analysis, and evaluation of results. Finally, conclusions are drawn from the results, limitations acknowledged, and areas for future research highlighted.

Chapter one – Literature review

1.1 Defining fatigue

Fatigue is a complex topic that can be easily misunderstood (Sharpe & Wilks, 2002). It involves a wide aspect of human capacity, including physical and cognitive activity, information processing, and emotional state. Furthermore, there is confusion surrounding classification of fatigue as different 'types' and classification by different causes. There is even confusion relating to the line between 'causes' of fatigue, fatigue itself, and 'symptoms' of fatigue; there is talk of the symptoms of fatigue, but for many medical sources, fatigue itself is referred to as a symptom, usually of a chronic disease or illness.

Fatigue, as a medical occurrence, can be either chronic or acute (Smith et al., 1999; Tan, Sugiura, & Gupta, 2002). Chronic fatigue is that which occurs repeatedly over the course of several weeks or months, and is usually caused by an underlying disease or medical condition. In contrast, acute fatigue is that which occurs over a very short timeframe, usually no more than several hours, and is often caused by factors such as lack of sleep, high workload, or other environmental stressors. Herein, the term 'fatigue' is used only to relate to acute fatigue, as chronic fatigue is of little importance in this study.

There are many problems associated with establishing a definition of fatigue, with some arguing that it is in fact a hypothetical construct (Bennett, 2003). Primary amongst these is the frequent classification of fatigue under three separate sub-groups which can be difficult to differentiate between: physical fatigue, mental fatigue, and emotional fatigue.

Physical fatigue is the decrease in physical performance and activity that can result from a number of varied causes; it often occurs after periods of extended or intensive physically-demanding activity, and simply put is the inability of muscles to generate their normal forces (Stokes & Kite, 1994). Apart from muscles becoming fatigued through use, other causes can include suboptimal nutrition, illness, and lack of sleep. Physical fatigue can be identified by the body being unable to complete physical tasks that it would otherwise be able to in a non-fatigued state. Other symptoms can include a feeling of soreness or tiredness.

Mental fatigue results from, as the name suggests, mental activity rather than physical activity (Wilson, Russell, & Caldwell, 2007). That is, an individual can suffer from mental fatigue without having to undergo any physically demanding activity. However, it is similar to physical fatigue in that the activity which causes the fatigue, while in this case mental and not physical, can be either extended or intensive. The primary symptom of mental fatigue is a decrease in mental performance, or information processing capacity (Lamond & Dawson, 1999). In addition to being caused by mental activity, mental fatigue can also be triggered by lack of sleep, suboptimal nutrition and illness in a similar way as physical fatigue.

Emotional fatigue can be identified as fatigue resulting from feelings or anxiety, worry or depression, and stress and pressure from outside sources (Trollip & Jensen, 1991). Similar to physical fatigue, the factors that cause emotional fatigue can be intensive, such as a heated disagreement, or extended, whereby an individual is worn down over a longer period of time due to anxiety or worry. Symptoms of emotional fatigue include decreased mental performance, and in this way it can be seen as being closer to mental fatigue than physical fatigue.

Despite the above types of fatigue seemingly being very different, it can be difficult to clearly differentiate between them, and boundaries between them can blur considerably, almost to the extent that in many cases a definite diagnosis of individual fatigue types can be impossible. For example, many causes of fatigue such as lack of sleep and poor nutrition can cause both physical and mental fatigue. In addition, some symptoms, especially those such as poor coordination which involve both a mental and a physical aspect, can be misinterpreted as either mental or physical fatigue. In any event, it is not even widely accepted that there is a distinct line between 'physical' and 'mental' aspects, the two being very closely linked. In conclusion, while in theory there are situations where fatigue can be simply classified as either mental, physical, or emotional fatigue, in practice this distinction cannot be made so simply.

With a classification of fatigue by type being potentially misleading at best, attempts have been made to classify fatigue by the factor that causes it. There have been considerable amounts of research into the effect of some kind of ‘fatigue’ on human performance (Stokes & Kite, 1994) taking place for over 100 years. In one of the first studies which took place in 1896, three volunteers remained awake for almost four days and performed various tasks at regular intervals (Patrick & Gilbert, as cited in Fuchs & Burgdorf, 2008). This study, and many others since, identified fatigue simply as degraded performance caused by lack of sleep, and it has been replicated many times throughout the 20th century such as in 1963 by Kleitman (as cited in Stokes & Kite, 1994). In aviation, similar studies have investigated the effects of lack of sleep, for example on pilots of F-117A stealth aircraft (Caldwell et al., 2003) and also civil airline pilots (Gundel et al., 1995). Indeed, it is this definition of fatigue associated with lack of sleep that is perhaps the most pervasive in the literature today, and while it is certainly not a complete assessment of the topic, lack of sleep is indeed one of the most common causes of fatigue in both aviation and more generic environments (Dawson & McCulloch, 2005; Ewing, 2003; Godwin, 2006). When viewed under the previous framework of fatigue ‘types’, this fatigue caused by lack of sleep would encompass both physical and mental fatigue, as it would result in both a decreased physical and mental performance.

In contrast, the classical view of fatigue as outlined by Bartlett in 1943, (as cited in Stokes & Kite, 1994) is simply degraded performance resulting from increased time spent on any particular task. While increased time spent on a task implies more time awake and therefore less sleep in the immediately preceding timeframe, this assessment was more concerned with the effects caused by the task itself, rather than the lack of sleep *per se*. For example, simply being awake but not undergoing any particularly strenuous activity can be seen as a ‘task’, even though it does not require any major effort. For example, after staying awake for twenty hours, the previous view would identify the lack of sleep as the cause of any fatigue, while a ‘time on task’ approach would identify the actual time spent awake, and the task [if any] carried out in this time, as the cause. A study fatigue in aviation safety by Goode (2003) inferred fatigue levels from time spent working, following the above approach of Bartlett, and found that the longer pilots spent working, the more likely an accident would be to occur. Of course, any activity which requires any kind of increased effort, whether physical or mental, can accelerate the onset of fatigue. The key is that fatigue is caused in direct proportion to the length of time spent on any one particular task. This view of fatigue would involve both mental and physical fatigue, because a 'task' could be either mental or physical. This kind of fatigue could be called 'exhaustion'.

These two approaches, 'lack of sleep', and 'time on task', while seemingly very different, can still relate and are not necessarily mutually exclusive. The 1896 study by Patrick and Gilbert focused on a lack of sleep, and hence an extended period of activity, while Bartlett's approach in 1943 was concerned with time spent conducting tasks, at the expense of sleep. Both studies suggest that the human body is limited in its ability to continue to perform any task at an acceptable level without adequate intervening periods of rest. An increased level of activity can be seen as artificially extending this 'awake' period by draining an individual's mental and physical resources quicker. Should this time without rest be extended, a reduced level of performance will be encountered.

With this confusion surrounding the understanding, evaluation, classification, and definition of fatigue it is not surprising that many suggest that there is no widely accepted definition of fatigue, even to the extent that each researcher simply creates their own definition (Bennett, 2003; Stokes & Kite, 1994). The lack of a widely accepted definition or definitions does not necessarily adversely affect or hinder research on the elusive topic, but it does raise challenges when attempting to compare different studies (Bourgeois-Bougrine, Carbon, Gounelle, Mollard, & Coblentz, 2003). Indeed, it does seem apparent the word 'fatigue' can mean different things to different people, and there does not appear to be any consensus amongst the literature (Bennett, 2003).

For example, considering the nature of this study, if a student pilot suffered from fatigue after a training flight, the way this 'fatigue' is defined can have a considerable impact on the results. The aim of this study is to investigate whether any fatigue results from completing a training flight. As flying is not a particularly physically intensive act, it is unlikely that any fatigue would be of the 'physical' type. There is the potential for emotional fatigue to play a part, as anxiety and externally- or internally-imposed stresses and pressure can occur in flight training. However, this would likely be an extraneous variable as it is not a direct result from the actual act of completing a training flight, i.e. manipulating aircraft controls, completing checklists, radio procedures, talking with the instructor, and other tasks that are requirements of a training flight. Furthermore, neither 'time on task' or 'lack of sleep' would be appropriate, as the training flights in question are usually limited to around one hour in duration. As a result, any fatigue resulting from the training flight would be more related to mental fatigue, as outlined above. So for this study a definition of cognitive fatigue, which will herein be referred to as simply 'fatigue' is as follows:

"That state characterised by a decrease in cognitive performance and information processing capacity resulting from the acute mental challenges placed on a student pilot from completing a training flight."

1.2. Causes of fatigue and complicating factors

Fatigue as a general concept can have many causes which can vary in their occurrence, intensity and frequency amongst individuals, making analysis difficult. While in this study there is only one cause being investigated, that is the "acute mental challenges placed on a student pilot directly resulting from and inherent to completing a training flight", there is the potential for other causes to create fatigue or fatigue-like symptoms as extraneous variables. Therefore it is important to understand and account for these other factors. Some of these key factors are summed up in the mnemonic IMSAFE [I'm safe], a variation of which is taught to student pilots as a means of self-evaluating their ability to fly safely. These six letters of IMSAFE stand for, in order: Illness, medication, stress, alcohol, fatigue [specifically sleep], and eating.

1.2.1. Lack of sleep

Of the many causes of fatigue, perhaps the most prevalent amongst the literature is lack of sleep (Caldwell et al., 2009; Lamond & Dawson, 1999). When considering the extent at which a certain reduction in sleep can be identified as a 'lack', the initial requirement would be to determine a 'normal' amount of sleep. An initial obstacle to this is the lack of consensus about an ideal or normal amount of sleep, usually measured in hours of sleep. While eight hours per day seems to be an often-quoted figure, this is often seen as varying highly, from as much as four to ten hours (Robson, 2008; Samkoff & Jacques, 1991). Of course, it seems logical that an ideal level will vary between individuals, but it can also vary depending on a number of criteria including the nature of any activity undertaken whilst awake, and other factors such as nutrition, illness or lack thereof, and any sleep debt (Ritter, 1993). Further complicating this is that while sleep is usually measured in a number of hours spent asleep, there is no guarantee that a certain number of hours will confer the same resting benefits between individuals, and even for the same person under different conditions and situations (Stokes & Kite, 1994). For example, an alcohol-induced sleep is generally not as restful per hour, compared to sleep initiated without the use of alcohol (Robson, 2008). All of these factors complicate the analysis of lack of sleep as a cause of fatigue. While it may be relatively easy to identify considerable lack of sleep, such as in the case of only three or four hours per day, it becomes much more challenging when considering an only moderate lack, such as in the range of five or six hours. If this is extended for a continued length of time, it may accumulate over time into a sleep debt having the same effect as a considerable lack of sleep over a shorter period (Samkoff & Jacques, 1991). While measuring sleep in hours does not accurately convey the amount of fatigue-mitigating rest, there does not seem to be another simple method of expressing an amount of sleep. When considering fatigue in a broader context, lack of sleep is an important concept, and as such, even if it is not likely to be an independent variable in fatigue research, such as for this study, it should still be monitored as an extraneous variable. It is highly likely that any fatigue measured may have resulted at least in part from a lack of sleep rather than any independent variables being monitored in the study.

Lack of sleep can also be caused by being awake at times when the body will naturally want to rest, these times being set by external stimuli called zeitgebers, meaning 'time givers' (Signal, Ratieta, & Gander, 2006). Zeitgebers include natural phenomena such as sunlight and air temperature, and other factors such as regular meal times and work shifts. They help to create a sleep/wake cycle, a pattern of times when the body is usually awake and asleep. Attempting to sleep out of sync with this cycle and zeitgebers can result in inadequate and ineffective sleep, which can lead to fatigue. Zeitgebers help to realign the body with a new timezone after rapid transit between different zones. However, the body's sleep/wake cycle maintains a level of inertia, and adjusting to a different time zone usually takes time. While this re-alignment takes place, the initial conflict between the sleep/wake cycle and zeitgebers can also lead to inadequate and ineffective sleep. These are common occurrences for long-haul airline pilots.

1.2.2. Illness and medication

Some illnesses and diseases cause fatigue, either directly as a symptom, or by increasing the likelihood of succumbing to fatigue through other causes such as lack of sleep. Fatigue can be caused by physiological diseases such as cancer, neurological diseases such as multiple sclerosis and Parkinson's disease, and psychological ailments such as depression. In these and other cases, fatigue "is often reported by patients as being amongst their most severe and distressing symptoms" (Dittner, Wessely, & Brown, 2004) However, many of these diseases are not relevant to aviation, with pilots and students being required to pass strict medical tests to screen out any potentially debilitating diseases and conditions. They are also required to inform their aviation doctor should any signs or symptoms of illness develop, and also not to fly should they suspect they are suffering from any kind of illness. Some comparatively minor illnesses such as Influenza can also have fatigue as a symptom, and this can persist after some of the more debilitating symptoms have passed. While it is a pilot's responsibility to ensure they are in a fit state to fly, it is possible that they may recommence flying too soon after being unwell, to the extent that they are still suffering from fatigue. Fatigue or fatigue-like symptoms can also be caused by medication being used to treat an illness. However, most of this type of medication is incompatible with flying because of side-effects on the brain and nervous system (Ewing, 2003).

1.2.3. Nutrition

Nutrition is an important concept when considering fatigue, as, in addition to sleep, it is a key component which allows the human body to continue to perform in an un-fatigued state, literally as "fuel for the body and mind" (Robson, 2008, pg 98). Regular and appropriate meals will allow the body and mind to operate at their optimum levels, and resist the onset of fatigue. The accident analysis of the crash of

Air New England [ANE] Flight 248 in 1979, a de Havilland Canada DHC-6-300 Twin Otter light twin-engined aircraft, contains some interesting information relating to nutrition and fatigue. At the time of the accident at 2248 local time, it could not be established that the captain had actually consumed any food of nutritional significance during the past 24 hours, the only recorded intake being "a Danish pastry and a cup of coffee in the late afternoon between flights" (National Transportation Safety Board [NTSB], 1980, p. 6). This was seen as highly likely to have led to the captain to be suffering from several debilitating symptoms at the time of the accident such as:

> "subtle mental confusion, slowing of cognitive processes, and diminution of psychomotor ability, [which] cannot be distinguished from symptoms of fatigue, and even when mild, would certainly contribute to the effects of fatigue. These physiological factors ... are known contributors to the degradation of human performance..." (NTSB, 1980, p. 16).

Robson (2008, p. 102) suggests that "many pilots fly partially incapacitated" from these symptoms. The captain was also taking unauthorised medication for hypertension that can possibly cause or accelerate the development of fatigue or fatigue-like symptoms. He had also worked a 14-hour duty day, which had been extended much to his annoyance; this emotional stress was well documented by eyewitnesses earlier in the day of the accident. The long day involved fifteen approaches and landings in the Twin Otter aircraft, "an aircraft well known for its dark, noisy, cockpit environment [that] would almost certainly have produced some ... fatigue even in the hardiest individual and particularly in [the captain]" (NTSB, 1980). The NTSB believes that these factors resulted in a significant degradation of the captain's physiological condition, which seriously impaired his performance. Other reports of pilots suffering from poor nutrition to the extent that it could have serious repercussions on flight safety are recorded in the NASA Aviation Safety Reporting System (Ritter, 1993; Stokes & Kite, 1994). While of course this is an extreme case, this accident does highlight the potentially disastrous effects of poor nutrition and its role in preventing fatigue. Somewhat related to nutrition is the use of caffeine as a stimulant to offset the effects of fatigue. While this can be achieved in the short-term, it is a negative health influence and can actually increase overall fatigue due to side effects of increased levels of stress, urination, headaches, and dehydration (Civil Aviation Authority [CAA], 2000; Robson, 2008).

1.2.4. Mental activity

Most pilots undergo varying degrees of mental activity which can range from intense single-pilot instrument flight rules [IFR] operations in a light twin-engined aircraft, to the comparative relaxation of a commercial flight crew monitoring aircraft systems en-route to their destination. Even within different roles, pilots often experience extremes of mental activity, usually reduced during the cruise phases of flight, and increased during take-off and landing. Heightened levels of mental activity, whether in quantity or duration, can lead to ‘mental’ fatigue, characterised by the feeling of exhaustion or being ‘worn out’, even though not having completed any physically demanding tasks. Indeed, a typical environment for a pilot to suffer fatigue from mental activity, such as in the single-pilot IFR example above, will involve the pilots strapped in their seat with restricted movement for several hours at a time. When mental activity is high enough, it may take little time for an individual to become mentally overwhelmed and suffer some degree of fatigue. This is especially relevant for student pilots who can find themselves bombarded by information from various sources in a new and possibly uncomfortable environment.

1.2.5. Physical activity

While excessive physical activity can cause fatigue, this is not often an obvious concern in aviation. Flying an aircraft, especially for commercial airline pilots, is most of the time not exactly a physically demanding occupation. While some pilots load bags and manhandle and refuel aircraft as part of their job, this does not usually require any particularly prolonged or strenuous effort. Of more concern is that pilots maintain an adequate level of physical activity in the form of exercise. It is highly recommended that pilots have a regular exercise routine with the aim of improving heart and lung fitness (Ewing, 2003). This also can help to prevent the onset of fatigue from other sources.

1.2.6 Stress

Stress is closely linked to fatigue, and is also very similar, being a varied and complex topic; fatigue has even been defined, somewhat simply, as an accumulation of stress

(Ewing, 2003). In fact, the concepts of stress and fatigue can overlap to the point of confusion, and much of what has been previously discussed about mental activity and mental fatigue could be interpreted in different terms, with respect to stress (Ewing, 2003). For the purposes of this discussion, stress will be investigated as an external demand on an individual, which may originate from many sources. There are two broad kinds of stressors: physiological or environmental, and emotional or psychological (Robson, 2008). External physiological stressors such as vibration, noise, and humidity and temperature extremes inherent in the aircraft environment can take their toll on a pilot in a similar way to strenuous physical exercise. Over time, their effects, while seemingly limited, can build and compound as outlined in the NTSB accident report of ANE Flight 248 discussed above under section 1.2.3. The accident report also commented on emotional or psychological stressors; these are also of considerable importance and relate to the topic which was briefly touched upon in section 1.1. Emotional stress can arise from many sources, from professional or personal issues, and can include financial concerns, employment uncertainty, death of a family member, divorce, disagreement with superiors or colleagues at work, a challenging work environment, pressure to continually perform at high levels, encouragement to 'bend' rules, medical difficulties, and many other issues (Bor, Field, & Scragg, 2002; Trollip & Jensen, 1991). Any of these stressors, such as the unsafe and unexpected prolonged duty day that the captain of ANE Flight 248 was required to work, can lead to emotional fatigue.

1.2.7. Causes of fatigue in different flight operations

Aviation is a complex and varied industry, with pilots flying many different aircraft types in different environments worldwide. These differing flight operations each create different potential causes for fatigue (Bourgeois-Bougrine et al., 2003), some of which will be highlighted below.

Long-haul international pilots typically become fatigued due to both a combination of Patrick and Gilbert's lack of sleep approach, and Bartlett's time on task approach. The lack of sleep quantity and quality can arise from disruption of the sleep/wake cycle due to time zone transitions, night shift work, and scheduling changes (Caldwell et al., 2009; Yen, Hsu, Yang, & Ho, 2009).

However, pilots flying regional and short-haul routes do not suffer from these same effects and instead become fatigued due to other causes, primarily the increased mental workload resulting from a higher number of takeoffs and landings per flight hour (Powell, Spencer, Holland, Broadbent, & Petrie, 2007; Yen et al., 2009). According to a New Zealand study carried out by Powell et al. (2007), commercial pilots flying 'short haul' operations of between 45 and 70 minutes several times a day are more likely to be subjected to fatigue than pilots flying 'long haul' international flights. Being rostered for a variable pattern of early starts and late finishes and the multiple take-offs and landings involved are the main causes of this fatigue (Powell et al., 2007). Pilots operating such high-frequency operations can become fatigued more quickly than those having to deal with the effects of long-haul travel. While the outcome as far as symptoms and effects can be similar, this fatigue is caused not by sleep disruption, but is rather cognitive fatigue caused by the demands placed on the pilot's information processing abilities. The concept of cognitive fatigue from high-frequency flight operations can be extended from short-haul airline pilots to any pilot who conducts similar operations, such as flight instruction, charter or scenic flights, or topdressing.

The causes of fatigue in the training environment are different from those that affect commercial pilots, especially long-haul international pilots. The first few training flights that student pilots undergo are in an unfamiliar environment of a light aircraft which can be cramped, noisy, and hot. They are also faced with the challenge of processing a vast array of unfamiliar information: interactions with the flight instructor, radio calls from air traffic control and other aircraft, information displayed on the aircraft's instruments, information from outside regarding navigation and other aircraft, all while trying to recall lesson details from a pre-flight briefing. It is easy to see how just one hour in this fast-paced environment which demands total concentration and features considerable mental challenges can lead to cognitive fatigue.

1.3. Effects and symptoms of fatigue

Fatigue has many symptoms, the occurrence and severity of each varying greatly between individuals. Some of these can include: decreased situational awareness, cognitive tunnelling, hazardous attitudes, diminished vision, reduced reactions, memory loss, poor concentration, increased likelihood of mistakes, diminished motor skills, substandard information processing capabilities, reduced communication abilities, and poor coordination (Caldwell et al., 2003; Fletcher, Lamond, van den Heuvel, & Dawson, 2003; Robson, 2008). It can be seen that the effects of fatigue are similar to that of excessive consumption of alcohol, and they are in no way conducive with safe and effective flying (Lamond & Dawson, 1999; Fletcher et al., 2003; Caldwell & Caldwell, 2003). A 1976 NASA study by Gartner and Murphy identified both a lack of consensus about the idea of fatigue but also suggested a link between high workload and fatigue, with the effects of both being similar. They identified four broad categories of the "unwanted" effects of fatigue in increasing levels of importance (Gartner & Murphy, 1976). The first category is motivation and includes negative feelings towards the environment, decreased morale, and hazardous attitudes (Ewing, 2003), but without affecting actual measurable performance. Reduced proficiency is the second category and includes deteriorating precision, coordination, timing, and reserve capacity to deal with unexpected events. The third category is psychological stress which involves impaired information processing and overall neurocognitive functioning; many of these effects and symptoms can be understood with reference to the human information processing system. Finally, the most severe level of unwanted effects involves outright errors.

1.3.1. Motivation and attitudes

Hazardous attitudes are frequently discussed in relation to aviation decision making, as individuals who adopt a hazardous attitude are at a higher risk of adopting faulty judgement and making improper decisions (Ewing, 2003; Hunter, 2005). Five major hazardous attitudes are usually identified: resignation, anti-authority, impulsivity, invulnerability, and macho (Hunter, 2005). Resignation is the tendency to avoid taking action and making decisions to change a situation, the belief that it is better to leave things to chance and that 'everything turns out ok in the end'. It may result in such situations as a pilot continuing with a marginal approach instead of making a go-around, or by pressing on into bad weather. Anti-authority is the rejection of rules and regulations imposed by superiors or governing authorities, or even the advice of others, and the belief that things can be done better without unnecessary restrictions. This could also lead to a pilot flying lower than allowed on a landing approach or proceeding when weather conditions fall below legal minimum levels. Impulsivity is taking action quickly, but without thinking through alternatives and considering different options available. In a critical safety situation it could result in major errors such as shutting down an incorrect engine in the event of a failure, such as in the crash of British Midland Flight 92. Invulnerability is thinking that “accidents don’t happen to me”, and can result in risk taking and allowing unsafe situations to develop unchecked. Finally, a macho attitude is one of superiority and showing off to impress others, even to the extent of taking risks and breaking rules.

1.3.2. Proficiency and skills

The degradation of proficiency and skills is the second most important, and the first measurable category of the unwanted effects of fatigue. Examples include reduced reactions, diminished motor skills and muscle memory, poor coordination, and impaired vision. Vision is perhaps the most important sense in aviation, and any kind of reduced vision capabilities can potentially be catastrophic, with consequences such as misreading crucial flight instruments, failing to spot conflicting traffic, and increased susceptibility to visual illusions (Ewing, 2003; Robson, 2008; Trollip & Jensen, 1991). Reactions, motor skills, muscle memory and coordination are obviously all important in the face of an adverse development such as an engine failure at take-off or a stall on approach, but there are other ways that poor reactions can be detrimental. For example, student pilots practicing manoeuvres such as aerodynamic stall recovery are often required to respond quickly to stimuli such as aircraft positioning, engine sounds, and instrument readings. The human information processing system has a measureable inherent delay in its functioning, but when this is increased due to fatigue, it can become more difficult to both complete training manoeuvres, and respond to real-life emergency situations. An example of this is the widely publicised crash of Colgan Air Flight 3407, a Bombardier Q400 which crashed on approach to Buffalo Niagara International Airport in 2009 (NTSB, 2010). The aircraft stalled at 2,000ft and the fatigued captain made an incorrect stall recovery which resulted in the aircraft crashing into a house below the flightpath. While there were other factors involved, the fundamentally erroneous recovery techniques employed by the captain were almost certainly linked to his fatigue levels.

1.3.3. Cognitive impairment

The third category of cognitive impairment covers many aspects of the human body's mental capacity. Short-term and long-term memory loss, decreased situational awareness, increased distraction, cognitive tunnelling, poor concentration and reduced information processing capacities are all symptoms that can result (Lamond & Dawson, 1999). Memory loss can include forgetting or incorrectly remembering data stored in the long-term memory such as memorised emergency checklists, to more recently acquired information such as a pre-flight briefing or air traffic control clearances. Situational awareness is a key aspect of aviation, and comprises three main steps: perception, comprehension, and projection; or more simply: gather, understand, and think ahead. It refers to the effective assimilation of information from all available sources, understanding what this information means about the current situation, and extrapolating it to the near future. Any breakdown of this process is a serious safety hazard. Distraction, poor concentration, and cognitive tunnelling are all linked to the characteristics and limitations of the human information processing system.

1.3.4. Human information processing

The human information processing system is the process whereby information is gathered from the senses, and transmitted to the central nervous system, comprised of the brain and spinal cord, to be processed and understood (Robson, 2008). The human brain is limited in its ability to process information, in that it receives multi-channel inputs but only has a single-channel output, which means that while information is being received from many sources, only one can be attended to at a time (Robson, 2008; Trollip & Jensen, 1991).

There are four levels of human information processing (Robson, 2008). The conscious level involves active intervention of the brain to make direct choices and actions, an example of which would be a pilot responding to an in-flight emergency such as a radio failure; active decision making is required to evaluate the problem and choose the correct action to be taken. The subconscious level refers to those learnt skills and procedures that do not require conscious input, an example being a qualified pilot maintaining straight-and-level flight in good VFR conditions - little active mental input is required. Autonomic processes are the third level, and include automatic physiological responses such as breathing and heartbeat. The fourth level of processing is reflex, referring to reactions which bypass the conscious level.

The brain initially deals with the multi-channel information it receives through the different senses by filtering what it perceives to be important and relevant information so that single-channel processing can take place (Robson, 2008). In an aviation context, this would mean visual information of sighting another aircraft nearby would take priority over information that is less important such as the pressure from the aircraft's seat and restraint against the pilot's body. All of the information is received but that which is seen to have a higher priority is forwarded to the working memory for further processing. The information that is judged to be of little consequence is briefly stored in the sensory memory for up to about five seconds before being lost (Campbell & Bagshaw, 1991). Without a filtering system, the working memory would be bombarded with a large quantity of data which would in turn exceed its processing capacity and result in normal information processing being impossible. When information from several sources needs to be processed simultaneously, priority can be shifted quickly back and forth between these sources. The filtering process can be adversely affected when compromised by fatigue, which can result in information being incorrectly filtered such that less important information is selected to attend to, and important information is ignored. This can be coupled with a reduced overall capacity to filter information. Filtering is conducted at the subconscious level, but it may sometimes be advantageous to consciously examine information from the senses to determine whether there are any incorrect filtering processes, such as in times of fatigue. However, this conscious examination takes valuable resources and deprives the conscious working memory of its ability to conduct other tasks.

After the brain has selected important information to be attended to by the working memory, it is required to be processed and understood. This process is called perception and involves comparing the information currently in the working memory with additional information stored in either the short-term or long-term memory (Campbell & Bagshaw, 1991; Robson, 2008). The sound of an engine running rough, for example does not necessarily mean anything by itself, but gains significance when compared to the sound of the same engine running normally as stored in the short-term [from just previously in the flight], and long-term [from past experience] memory. For example, when hearing a radio call, past experience can help a pilot to ‘fill in the blanks’ and easily make sense of what might otherwise be an incomplete or ambiguous transmission.

An experienced pilot can split their workload between conscious and subconscious processing, by allowing basic flying skills and routines to be completed subconsciously, with only minor occasional conscious input when required. This leaves the majority of the conscious processing capacity to deal with other more important tasks, such as unanticipated situations that may arise during the flight (Robson, 2008). Fatigue from information processing overload is particularly important to student pilots. In contrast to a qualified and experienced pilot, a student pilot does not have the ability to allocate many, if any, tasks to the subconscious processing level. This results in the conscious processing capacity being required to work hard to attempt to deal with all the unfamiliar information that it is receiving. Some of the many effects of fatigue such as slowed reactions, memory problems, impaired vision and motor skills, increased likelihood of mistakes, and decreased situational awareness have the potential to degrade a student's ability to effectively complete their flight training (Campbell & Bagshaw, 1991).

When an individual is affected by fatigue, they are less capable of efficiently and effectively processing information. This can result in information being improperly prioritised by paying attention to trivial matters at the expense of more important matters. An extreme version of this is cognitive tunnelling, which occurs when one particular task is given a very high priority at the expense of other tasks (Crawford & Neal, 2006). It can be especially relevant when the task being focused on is actually less important than those tasks being neglected. The crash of Eastern Air Lines Flight 401, a Lockheed L-1011-1 TriStar wide-bodied airliner, highlights the potential effects of cognitive tunnelling (NTSB, 1973). The flight crew became distracted by trying to replace a lightbulb in the landing gear indicators; cockpit voice recordings indicate the pilots likely experienced cognitive tunnelling as they tried to establish a way of replacing the lightbulb without breaking its cover. In this case, all of their attention was given to this one small problem, at the expense of flying the aircraft. The aircraft began slowly descending over a period of about five minutes until it collided with the terrain below, and at least four obvious indications that this was taking place went completely unobserved by the crew (NTSB, 1973). The human information processing system is more susceptible to these types of mistakes when affected by fatigue.

1.3.5. Human error

Finally, all these effects of fatigue culminate in the most serious category, that of outright human error. Error is often sorted according to two broad categories, an incorrect act known as a mistake, and a correct but improperly executed act known as a slip or a lapse (Reason, 2000). Any error is unsatisfactory by its very nature, and especially so in aviation where it can often lead to critical safety failures (Campbell & Bagshaw, 1991).

1.3.6. Fatigue in aviation safety

In the aftermath of a serious aviation incident or accident, considerable effort is usually put in to investigating the factors that contributed to the event. As a result, investigation reports are a valuable source of information of the potential effects that fatigue can have on human performance.

Aviation safety is a highly complex endeavour. While attempts are continually made to pre-empt safety failings through research, and analysis of incidents, unfortunately it is often major accidents that provide important information and act as the catalyst for changes. When considering fatigue, immediate problems arise when attempting to quantify its effects on aviation safety. In the aftermath of an accident, investigators often have very little information to suggest whether the key people involved such as pilots and air traffic controllers were suffering from fatigue, and if so, whether it had an effect on the accident. In the United States, fatigue is a causal factor in approximately 20% of accidents and incidents across all sectors of flight operations (CAA, 2000; Jackson & Earl, 2006). Comparatively, in New Zealand the CAA's accident records show that only 0.2% of accidents and incidents have fatigue as a causal factor (CAA, 2000). However it is not very likely that these records reflect its true effects. Safety experts within the CAA suggest the real figure would be much higher, at approximately 25%, and that from what is known of other countries, it is almost certain that "fatigue is chronically under-identified" (CAA, 2000). This could potentially be due to under-reporting by individuals involved, lack of understanding, and the inherent difficulty in identifying and quantifying fatigue.

Fatigue continues to play a major role in aviation safety, and it has been suggested that the majority of errors made by pilots are of a cognitive nature, influenced by fatigue (Ritter, 1993). The negative consequences of fatigue have played a key part in many aircraft accidents, one such example being Korean Air flight 801, a Boeing 747-300 which crashed on final approach to Antonio B. Won Pat International Airport in Guam on August 6, 1997. At about 0139 local time in the midst of heavy rain and cloud, flight 801 was cleared for a difficult non-precision localiser-only Instrument Landing System [ILS] approach to runway 6L (NTSB, 2000). In the midst of the approach a conversation recorded on the cockpit voice recorder [CVR] indicated that the flight crew began a discussion about the approach guidance systems and the aircraft began a steep descent, prematurely penetrating two descent altitudes. As the aircraft approached a hill below, the ground proximity warning system [GPWS] sounded an alert, and in response the first officer said "let's make a missed approach". However, no decisive action was taken and 13 seconds later the aircraft crashed into the side of a hill, still 4 miles from the runway threshold.

Subsequent investigations determined that the major causal factor was the crew's actions, most likely brought on by fatigue (Caldwell, 2005). The captain's fatigue level was likely high, as the only reason he was flying that particular route was because he had insufficient duty time to fly his scheduled route to Dubai. This indicated he was operating close to his rest limit and therefore quite possibly fatigued. The captain made several comments about working conditions that were picked up on the CVR, including: "probably this way, hotel expenses will be saved for cabin crews, and maximise the flight hours. Anyway, they make us [747] classic [pilots] work to maximum", and later, "eh…really…sleepy" (NTSB, 2000). As the effects of fatigue are often underestimated by individuals, the fact that the captain made these unsolicited comments likely indicated a significant level of fatigue (Caldwell & Caldwell, 2003). Furthermore, at the time of the accident, 0142 local time, it would have been 0042 in the captain's home time zone, a time when he would usually be sleeping. While accident investigation can be an imprecise science, some of the effects of fatigue can be matched up to errors likely made by the flight crew. Memory loss caused the initial confusion over the approach guidance system despite clear instructions from air traffic control. Reduced information processing capabilities likely lead to cognitive tunnelling, which lead the pilots to focus on the approach guidance at the expense of actually flying the aircraft. It also led to decreased situational awareness, meaning the pilots did not recognise their dangerous descent and proximity to the terrain below, while reduced vision prevented the pilots from noticing their decent as shown on cockpit displays. Finally, reduced reactions and poor decision making meant a go-around was not initiated in time.

As can be seen, the effects of fatigue fit with errors committed by the flight crew in the accident sequence. Even though the flight crew had clear instructions from air traffic control they spent a large amount of time in the critical approach phase discussing the approach guidance system. Aside from the usual approach checklist items such as landing gear and flaps, conversation until the GPWS alert was entirely focused on the approach guidance systems. It is likely that the crew, and especially the captain, experienced cognitive tunnelling while addressing this comparatively unimportant subject, when instead they should have been concentrating on the non-precision approach.

Few aviation incidents or accidents are caused by a single factor. A widely accepted description of the events of any aircraft accident, indeed almost any accident, is the "Swiss cheese model" proposed by Reason (2000). The main idea of the Swiss cheese model is that a number of active and latent failures combine to overcome defence layers in place to prevent accidents. These defence layers are seen as stacked slices of Swiss cheese, with the ubiquitous holes viewed as system deficiencies which prevent the defence layer from acting effectively. Defence layers can include safety equipment, procedures and regulations, assistance from air traffic control, and finally the pilot, who is often seen as the last layer of defence against accidents. Holes in these layers could be caused by any number of things including faulty, inoperative, or missing safety equipment, inadequate regulation, ineffective operating procedures, poor safety culture, and suboptimum pilot performance. In order for an accident to occur, every layer that is in place to prevent such an accident must be breached, and this requires a single catastrophic failing, or more frequently, a number of causal factors. An alert, non-fatigued pilot can identify dangerous situations and prevent accidents even if other earlier layers are breached, but if a pilot is fatigued and hence operating at a suboptimal level, holes begin to appear in this last line of defence.

Another way of looking at causal factors in aircraft accidents is as an "Accident causation chain". The failures can be thought of as a chain of events with each individual factor forming a link in the chain, which when complete, results in an accident (Campbell & Bagshaw, 1991; Miyagi, 2005). The defence layers in the previous Swiss cheese example are instead viewed as possibilities to break the chain by destroying individual links. A chain that is forming through a number of factors can be broken by an alert and non-fatigued pilot. However, should the pilot be fatigued and not operating at an optimal level, it is possible that potential opportunities to break the chain can be missed and an accident becomes inevitable.

1.4. Measuring fatigue

Several problems become apparent when attempting to measure fatigue, mostly relating to the lack of certainty and consensus about fatigue and its causes and effects. Fatigue can be caused by many factors including physical exertion, lack of sleep, and mental challenges; these different causes can lead to fatigue with different effects. The occurrence and strength of different effects can further vary between individuals in similar situations. Because of the many extraneous variables that can affect fatigue and create similar effects it can be difficult to distinguish between symptoms resulting from fatigue, and those occurring from other sources. The effects themselves can also be difficult to objectively measure and quantify, with complicated phenomena such as cognitive tunnelling and situational awareness being almost impossible to accurately measure; it is usually only possible to measure some simpler factors. Finally, when individuals are asked to evaluate their fatigue, they almost always underestimate its effects (Montgomery, 2007; Stokes & Kite, 1994).

1.5. Research overview

The literature review clearly outlines the many negative effects that fatigue can have on human performance. These effects are even more important in the aviation environment, as it requires high performance from individuals at all times to ensure efficient and safe operations. The many studies that have focused on fatigue in aviation have found that it can be a considerable impediment to safety and efficiency, and investigations have shown that it has been a causal factor in many aircraft accidents. Several aspects of aviation have been assessed in relation to fatigue, from commercial to military pilots, and air traffic controllers. Many of these studies and investigations have been prompted by the major consequences that can result from errors caused by fatigue. Many lessons learned from commercial and military aviation have been applied to sections in the GA environment, such as flight instructors and charter pilots. However, flight training has, to date, been an area that has received little attention. Possibly a major reason for this is that student pilots do not suffer from many of the usual causes of fatigue that pilots in other environments are subject to. A small number of short day-time flights each day are unlikely to present many of the more common causes of fatigue. In addition, the presence of a flight instructor on many training flights acts as an additional layer of safety should the student's performance be impaired by any fatigue. Cognitive fatigue can be caused by mental workload however, and there are few situations in aviation where mental workload is higher than in flight training. However, the negative effects of fatigue on human performance relate not just to safety but in a broader sense, the ability of a pilot to complete the objectives of a flight. In the case of commercial aviation, this is usually simply the safe, timely, and comfortable trip between two locations. However for a student pilot, an additional objective of effective learning and development of skills is an important additional consideration. While the level of learning and development relies on an effective instructor, this is of little use if the student is fatigued enough that their information processing capability is sufficiently reduced to make effective learning impossible. Effective learning is very important both for individual students, flight schools, and the aviation industry as a whole.

With little research in the area of fatigue in flight training, an initial step would be to attempt to analyse fatigue levels in student pilots so as to set a baseline from which future study can be built on. This study attempts to set such a baseline through exploratory research involving student pilots. Students in the early stages of their training are less familiar with the flight training environment and learning new techniques, and are therefore more likely to suffer from overload as a result of processing information at the conscious level. They are likely to be more affected by fatigue than those students completing more advanced flight training, and it would therefore be easier to identify fatigue in these individuals.

In light of the lack of previous research on the subject, two tasks were identified which needed to be completed initially in order to successfully complete the main body of the research. The following three chapters will address these two preliminary tasks, and the main study in order. Chapter 2 researches the population of student pilots training in New Zealand, which will then provide guidance about the sample to be used for the main research. Chapter 3 covers the design of two tools to measure fatigue, and a pilot study for validating these tools, which will then be utilised in the main research. Chapter 4 details the procedures and results of the main study.

Chapter two – Population

2.1. Introduction

In New Zealand, the requirements for obtaining a pilot licence are outlined in Civil Aviation Rule [CAR] Part 61 which is promulgated by the Civil Aviation Authority of New Zealand [CAA] (CAA, n.d.). There are two main streams of learning required: a series of six theory papers, and a programme of practical flight instruction. In order to obtain a Private Pilot Licence [PPL], which is the first step in any flight training, students are required to complete a legal minimum of 50 hours of total flight time [TFT], although in practice this number is usually closer to 60 or 70 hours depending on the structure, if any, of the flight training course undertaken, and other operational factors. The practical flight instruction is normally completed in a small GA aircraft, usually featuring either two or four seats and in the approximate range of 750kg to 1,100kg maximum certificated take-off weight. These aircraft are often owned by a flight school or flying club, and likely to be in demand from other students and pilots. In addition, as practical instruction, including aircraft hire and instructor's fees, is usually charged on an hourly basis, there is often a certain amount of pressure on the student to complete the practical instruction as soon as possible, with regards to both the number of flight hours and the time taken to complete the training.

The majority of initial flight instruction is completed with a certified flight instructor sitting in the front right hand seat, where the first officer or co-pilot would usually sit on a non-training flight, while the student sits in the front left hand seat, where the captain or pilot-in-command would usually sit. The instructor can fly the aircraft, demonstrate manoeuvres, and talk to the student as required, by virtue of dual linked controls required in all training aircraft cockpits. This kind of flight instruction is known as 'dual' instruction, also known as a dual flight. The process of a typical dual flight would involve the student completing pre-flight checks of the aircraft to be used for the training flight, initially under the direct supervision of the instructor, followed by a pre-flight briefing, then the flight itself, and then concluding with a post-flight briefing.

For a student pilot, this dual instruction can be somewhat intense as being in a training aircraft is often a new and unfamiliar environment, and they are also forced to deal with several new external stimuli. There are many visual aspects including the cockpit instruments and displays inside the aircraft, parts of the aircraft such as engine canopy and wingtips external to the cockpit, other aircraft flying nearby, natural and man-made terrain features, and weather phenomena. These many visual aspects require the student to actively manage their time and attention between them as required (Robson, 2008). Audible stimuli comes from three main sources: radio conversations over the aircraft's intercom from ground-based and air-based sources which may be directed at the student or between other third parties, interaction with the instructor, and sounds from the aircraft itself such as warning buzzers and engine sounds. Finally, further stimuli include vestibular, tactile, and ergonomic considerations such as loading or g-force, turbulence, aircraft control positioning and feedback, and the cockpit environment which can feature temperature extremes, vibration, and movement restrictions from its often cramped nature coupled with harnesses or restraints.

In addition to these stimuli, students may also be expected to recall points from theory papers and pre-flight briefings, recall and complete in-flight checklists, as well as recording other information. They are also required to physically manipulate the aircraft's many controls, often in a nature that is more challenging than a normal non-training flight. It is possible for flight instructors to gradually ease a student into many of these tasks and shield them from some of these aspects of flight training by assuming more responsibility for the conduct of the flight. However, the student must become familiar with these aspects as soon as possible, and so there is a good chance that keeping up with these many factors would present a difficult task to a student who is not familiar with the environment of a training aircraft.

In New Zealand, flight schools can be certified by the Civil Aviation Authority [CAA] under Part 141 of the Civil Aviation Act to operate as Certified Aviation Training Organisations [herein referred to as ATOs]. These ATOs must adhere to Part 141 requirements while conducting their flight training (CAA, n.d.). The CAA maintains records of all such ATOs, and as such it is relatively easy to obtain information about them. Most ATOs operate in a more formal 'school' environment, where students fly in a constant environment from day to day and flights are scheduled at semi-regular times.

Flight training is possible outside of an ATO, without the requirement to follow Part 141. This training is usually conducted at small aero clubs or flying clubs which, while required to use certified flight instructors to train students, often operate on a more infrequent and informal basis; however, there are larger flight schools that operate in the manner of an ATO but are not certified under Part 141. Herein all such 'uncertified' organisations will be referred to as 'flying clubs' for ease of reference. A number of student pilots in New Zealand train at these non-certified flying clubs, but obtaining information about them and their students is comparatively more difficult as there are no records kept by the CAA. Not operating under Part 141 also places fewer requirements on the flying clubs to maintain records of student numbers and training progress.

At the time of the commencement of the research, there was little readily available specific information about the number of student pilots training in New Zealand. Therefore, an initial objective was to investigate the population and gather information about both student numbers and training locations. It was important to understand the characteristics and distribution of the population before proceeding with any further research, as this information would influence the decisions made regarding the criteria for selecting a sample, and further design of the main study. The research problem was that there was no easily accessible information about the population. From this, two research questions were identified:

1) What are the numbers and distribution of student pilots training towards a fixed-wing pilots licence in New Zealand?

2) In light of the population data, what kind of sample allows for the most effective and efficient research?

2.2. Methodology

The primary method used to gather information regarding the population was direct contact with the CAA. Details are provided on the CAA's website of all "Part 141 Training Organisation Holders" (CAA, n.d.), and further communication with CAA licencing and flight training staff revealed that no records are kept of training organisations and flying clubs who do not operate under Part 141. From the 55 ATO's listed on the CAA website, 16 offered comprehensive 'flight school type' fixed-wing pilot training at the time of enquiry. Based on this information, direct inquiries were made to all 16 of these ATOs; it was established that student numbers vary considerably even from month to month as students complete their training, often at differing rates, and new entrants arrive, in up to several intakes per year.

Regarding uncertified training, Flying New Zealand [Flying NZ] is the "umbrella organisation" for the majority of fixed-wing flying clubs within New Zealand, which represents the majority of non-certified pilot training (Flying New Zealand, n.d.). However, there are still flying clubs that are not affiliated with Flying NZ, one example being the Air New Zealand Flying Club based in Christchurch which has a fleet of four Cessna aircraft and can train small numbers of pilots on an on-demand basis.

Because flying clubs often lack the structured courses and regular intakes that characterise ATOs, they usually do not keep records of student pilot numbers. In fact the line between student and pilot can often be somewhat blurred, with pilot members holding a licence undergoing further training to obtain type ratings, endorsements, and to further their flying skills. Newer ab-initio students training at flying clubs can also often vary in their characteristics, even at similar stages in training. Students of these flying clubs are more likely to train at infrequent intervals, and often with complicating factors; factors such as work and family life are a few of many which may potentially interfere with training and serve as extraneous variables in investigating fatigue.

When considering the direction of this study in relation to the nature of New Zealand flight training, especially between ATOs and flying clubs, there is considerable reason to believe that focusing on only certified ATOs is the optimal choice. Certified ATOs are easily identified and contacted, with access to student numbers, both current and projected, and stages of training. Students are also very similar in their characteristics and training programmes, with fewer extraneous variables to complicate data analysis, and conducting flights at regular times also makes data gathering more efficient. As a result, it was decided to only focus on ATOs in the interest of efficiency and effectiveness of the study, albeit at the expense of neglecting an important sector of flight training.

2.3. Results

At the time that inquiries were made in late 2012 there were approximately 800 students undergoing both full-time and part-time flight training at these 16 Part 141 fixed-wing training organisations.

Part 141 accredited ATO	N of students
Air Gisborne	4
Air Hawkes Bay	73
Air Napier	5
Auckland Aero Club	46
Canterbury Aero Club	80
CTC Aviation Training (NZ)	250
Eagle Flight Training	27
Kapiti Districts Aero Club	5
Mainland Air Services	24
Massey University	110
Nelson Aviation College	60
New Plymouth Aero Club	11
North Shore Aero Club	24
Southern Wings	38
Waikato Aero Club	22
Wakatipu Aero Club	20
Total	**799**

Table 1: Number of ATO students

It is important to note that these figures represent student numbers as of late 2012, and would have likely changed since then due to the previously mentioned fluctuation.

2.4. Discussion and conclusions

Because no reliable and efficient way was available to gather information on student numbers at non-certified flying clubs and it was unlikely to play an important role in the study, this was not attempted. Because of this, the exact population size of all fixed-wing student pilots cannot be accurately identified. However, it was identified that approximately 800 student pilots are at various stages of completing fixed-wing flight training courses at 16 ATOs across New Zealand, with the majority of these students located at 12 larger ATOs.

Out of the 16 CAA certified ATOs identified previously in chapter two, there are only twelve with student numbers high enough to warrant participation in the study. Due to the nature of the testing involved in the study, it is necessary for the researcher to be present on-site to conduct the testing. Travelling to a distant ATO may well enhance the reliability of the study, but with the potential of delays caused by weather turning such travelling into an expensive exercise it was elected to instead focus all available attention on a single local ATO. Possible disadvantages of focusing on one ATO include limiting the sample of participants from the single ATO and hence possibly affecting any possible generalisation of results. However, it does not seem likely that students at one particular ATO should differ in any significant way from those training at other locations. The twelve larger fixed-wing ATOs mostly offer a similar level of training in similar environments. CAA training requirements regarding curriculum, instruction, and examination ensure that all pilots are trained to the same standards.

Chapter three – Design and validation of tools

3.1. Introduction

In order to measure the fatigue levels of participants, it was necessary to design tools to accomplish this task as a search of the literature did not uncover any such pre-existing tools that could be used. Bourgeois-Bougrine et al. (2003, p. 1072) suggest that "the major problem with fatigue issues is the lack of … a reliable and valid assessment tool to measure it". Similarly, Goode's study (2003, p. 4) of aircraft accidents also noted "that there are no direct measures of fatigue or its onset". It was decided that two tools would be designed to measure fatigue, a questionnaire and a computer-based reaction time test, in order to gather as much accurate data as possible. Decreased reaction times are a frequently-observed symptom of fatigue, and they are easily measured and quantified (Caldwell et al., 2003). However, upon completing the tools it would not be known if they would produce accurate and reliable data. Therefore, it was necessary to complete a pilot study with a smaller, separate population, with the aim of validating the tools so that they could then be used to gather information from participants in the main study.

While fatigue has many varying symptoms, it was decided that the questionnaire would focus on only a few factors that participants could easily recognise, assess, and measure, to ensure accurate responses. The questionnaire was split into two parts, a pre-flight section that was administered before the training flight, and a post-flight section that was administered after the training flight. Due to the regularity with which individuals consistently under-evaluate their fatigue, it was decided to design the questionnaire in such a way as to create low face validity. As a result, the use of words or phrases relating to fatigue was avoided. However, it was necessary not to go to such an extreme as to make it deceptive or misleading as this would breach the requirements of the Massey University Ethics Committee.

In order to assess fatigue as accurately as possible, and to further avoid the well-documented regularity with which individuals underestimate their own fatigue, an objective reaction time test was incorporated into the study. It was based on the Psychomotor Vigilance Task [PVT], also known as Psychomotor Vigilance Test, a commonly-used tool to measure fatigue, particularly fatigue resulting from sleep loss (Petrilli, Roach, Dawson, & Lamond, 2006; Wilkinson & Houghton, 1982; Wilson, Russell, & Caldwell, 2006). The original PVT design measures participants' response times to a visual stimulus over a ten minute period. This is done through the use of a hand-held device which displays visual stimulus at varying intervals, and requires the participant to quickly press a button in response to the visual stimulus; the participant's reaction time is then recorded. The PVT has been used in many studies and evaluated as a reliable, accurate, and objective method of measuring fatigue. There are several different variations of the PVT in use; because of its simple design it has a very flexible application, and can be operated from a variety of electronic devices including purpose-built hand-held devices, computers, and smartphones.

In designing and validating tools, two main objectives were identified:

1) How can tools be designed in order to effectively measure fatigue in the challenging environment of flight training?

2) How can these tools be evaluated in order to ensure their reliability and validity?

3.2. Methodology

3.2.1. Questionnaire

The core part of the questionnaire was a ten-item measurement of fatigue. These items were assembled from and influenced by previous studies of fatigue in medical environments, particularly the 'Fatigue Assessment Scale' (Michielsen, de Vries, & van Heck, 2003; Michielsen, de Vries, van Heck, van de Vijver, & Sijtsma, 2004) and the '14-item Fatigue Scale' of Chalder et al. (1993). Other items were influenced by further studies (Beurskens et al., 2000; Lee, Hicks, & Nino-Murcia, 1993; Schwartz, Jandorf, & Krupp, 1993). While none of these studies specifically focused on fatigue in aviation, it was found that several items could be used or adapted to be made appropriate for an aviation environment. Initially, 22 items were identified that had the potential to be used, which then were narrowed down to 10 particular items that represented a condensed and concise, yet complete, overview of the 22 initial items. The items were worded so that they could be responded to with a seven-point Likert scale, ranging from 1 = Strongly Disagree to 7 = Strongly Agree.

These ten core items were then compiled into a pilot questionnaire which was administered to willing participants in several convenience samples, including individuals before and after a day's strenuous, but not abnormally so, activity, and shift workers before the commencement of an early morning start at about 4:00am. The aim of this pilot study was to serve as a validation study, specifically, to assess the internal consistency of the ten items so they could be split into two groups of five items between the pre-flight and post-flight sections.

3.2.2. Reaction time testing

Loh, Lamond, Dorrian, Roach and Dawson (2004) identify aviation as an environment where the ten minute length of a standard PVT is not ideal due to operational requirements preventing a test for this length of time. For example, studies of commercial airline pilots have found that PVT use was greatly restricted due to operational demands (Petrilli, et al., 2006; Rosekind et al., 1994). Operational demands can also present in flight training, such as at the ATO where the testing would take place, where students are required to maintain on-time performance [OTP], which is essentially the need to keep to scheduled flight times as much as possible. This restricts the length of time available for testing, at the risk that the participating students' OTPs suffer from the potentially 30 minutes required to complete the testing should ten minute PVTs be used. The testing should ideally be completed as soon as possible before and after the training flight to control extraneous variables and maintain consistency. As a result, it was decided to use a five minute version of the PVT, whereby the exact same test was used, but for only half the time. Several studies have shown that while a ten minute PVT may be ideal, for situations where time restrictions require a shorter test a five minute PVT can be just as effective and provides a viable replacement (Petrilli et al., 2006; Loh et al., 2004). Indeed, a five minute PVT called Reaction Self Test has been used on board the International Space Station to measure crew fatigue since Expedition 21 in 2009, and a version called PVT Self Test developed for a simulated 520-day spaceflight mission (Dinges, 2012; National Space Biomedical Research Institute, 2010).

The Psychology Experiment Building Language [PEBL] is free-to-use software which can be used to create and perform a variety of different psychological experiments (Mueller, 2010). One such experiment is the PEBL Perceptual Vigilance Task [PPVT], a version of the PVT. Based off a computer, the participant is required to depress the space bar in response to a red circle which appears in the middle of a black background after varying intervals between two and twelve seconds. The PPVT can be designed to finish after five minutes of testing and individual reaction times to each stimulus are then recorded. This study made use of the PPVT as the objective test.

3.3. Results

The participants reported no difficulty in completing the pilot questionnaire, and all of the items were easily understood, without any clarification required. However, in the test of shift workers, several returned questionnaires included missed items, and some even included contradictory responses, for example, strongly agreeing with both phrases: “Mentally, I feel in good condition”, and “I feel mentally exhausted”. Such a response was an indication of not fully understanding the items or the answer format. This suggested that care should be taken when designing and ordering the items to avoid any confusion, and that the questionnaires should be checked for any missing responses prior to completion. Cronbach’s alpha was chosen as the primary means of evaluating internal consistency. The results of the pilot study questionnaire were analysed and returned a Cronbach’s alpha of 0.79, which suggests high internal consistency.

A brief pilot test of the PPVT showed that the software performed as expected, and accurately measured the response time of participants. Participants found the test easy and intuitive to use, and no problems were identified.

3.4. Discussion and conclusions

The result of this favourable test meant that it was possible to split the ten core items into two groups of five, with one group to be used in the pre-flight section and the other group in the post-flight section. This allowed the two questionnaire sections to measure fatigue in similar ways, but using different items. If the same items were merely repeated in the pre-flight and post-flight sections, it would be possible for participants to be influenced by remembering their answers to the items in the pre-flight section and answering the post-flight section based on this recollection.

Of the ten items, five had a wording which related to a negative effect of fatigue such as “I feel fatigued”, or “I feel worn out” whereby the higher the level of agreement, or the higher number selected on the Likert scale, the higher the level of fatigue indicated. The other five items had a wording such as “I feel mentally alert”, which related to a condition which would not be expected from a fatigued individual. In this case, the higher the level of agreement, the lower the level of fatigue indicated. In the interest of maintaining two balanced sections it was decided to split these different items as evenly as possible; specifically, two of the former ‘fatigued’ items and three of the latter ‘non-fatigued’ items to make up the pre-flight questionnaire fatigue scale [QFSpre] and three ‘fatigue’ and two ‘non-fatigue items in the post-flight questionnaire fatigue scale [QFSpost].

The questionnaire featured a cover page with information for the participants including an ethics disclaimer and details on participants’ rights and the study’s privacy policy [See appendix A]. In response to the problems identified in the pilot study, clear instructions were given to the participants on completing the parts of the questionnaire that required Likert responses, and a reminder to check that it had been fully completed before submission. The first section gathered basic information about participants so that the information could be more accurately assessed. This included demographic information, specifically age and gender, flying background, details about the upcoming flight, and information on possible confounding variables, namely sleep, caffeine, and food consumption. The second section of the questionnaire was a list of items which required a response using a seven-point

Likert scale. The five pre-flight core items were embedded in this list, along with three other items more directly relating to the upcoming flight.

The post-flight section of the questionnaire featured a similar overall design to the pre-flight section, and comprised twenty items which required a response using a seven-point Likert scale. The five post-flight core items were included amongst items relating to the participant's analysis of the outcome of the flight, and their performance throughout it.

Chapter four – Main study

4.1. Introduction

The aim of the final major study is to measure and evaluate, and set a baseline for cognitive fatigue accumulated by student pilots over the course of a one-hour training flight. It is an exploratory study which will collect both subjective and objective quantitative data from a small sample of student pilots, building upon information gathered in the previous two sections in order to achieve these tasks. First, with the information on the population which was gathered in section two, decisions can be made on the type of sample to be used. Then the design and successful validation of the two tools will be further detailed, which allows testing to proceed. Finally, the testing procedure is outlined, followed by an overview of the data gathered. Six main research questions were identified, stemming from the overall research problem that there is no data about how student pilots are susceptible to fatigue:

1) Do student pilots demonstrate any fatigue as measured by a subjective tool?

2) Do student pilots demonstrate any fatigue as measured by an objective tool?

3) Do the results of the subjective and objective tool correlate?

4) Are the results statistically significant?

5) Are the results practically significant?

6) Can the results be used to set a baseline for cognitive fatigue?

4.2. Methodology

4.2.1. Sample

A large Part-141 certified ATO was approached to participate in the study. After the study was approved, participants were selected in a convenience sample from both direct communication with students and through notices distributed at the ATO premises. Due to low numbers of participants it was not possible to use a probability sample. The participants were tested as they expressed interest and became available at an appropriate time.

The sample size was limited by the number of willing students in the current intake, and the speed at which testing could be completed while the students were still completing training at stages that were appropriate for testing.

Due to factors such as weather, aircraft availability, and other operational factors, it was experienced early in the testing process that students could complete the pre-flight portion of the testing but then not end up completing their flight. In these situations it was decided to have the participants complete the 'post-flight' testing as much as possible after an appropriate amount of time had elapsed. Certain questionnaire items were therefore not applicable. This effectively created a small pseudo 'control' group, five participants who completed some parts of both the pre-flight and post-flight testing but without an intermediary training flight. While this control group was not an intention at the commencement of testing, but rather arose throughout the testing process, it did provide some interesting data for subsequent analysis. It was not possible to re-test these control participants on a future flight as this would result in discrepancies with some participants having then completed testing twice. Due to the relatively limited number of participants available, it was not possible to extend the control group to an equal number of participants as the main group as doing so would have necessitated deliberately sacrificing some of the participants from the main group. The final sample eventually comprised twenty-one participants undergoing a full-time flight training course at the ATO, 17 of these participants made up the main group, while the remaining 5 participants comprised the control group [See Table 1].

Of the seventeen participants in the main group, the average age was 23 years old, with a range from 18 to 36, and standard deviation of 4.4 years. However, there was an outlier, a 36-year-old participant who significantly skewed the data. Without this participant, the previous figures become an average age of 22, with a range from 18 to 26, and standard deviation of 2.69. There were only two female participants, both of whom were in the main group. The average total flight time of the participants in the main group was 69 hours. Recent sleep time in the preceding 24 hours [RSleep], at an average of 7.4 hours, very closely matched with the usual hours of sleep indicated [USleep], at an average of 7.6 hours. In the main group, six hours was the lowest number of hours recorded for sleep in the preceding night.

Aircraft types varied between participants, including several Piper models: 'Pa 28-112 Tomahawk', 'Pa 28-161 Warrior', and 'Pa 28-181 Archer'. Cessna 172 and Alpha 160 were the other two types used. Five of these flights, those completed with the Warrior and Archer models, were aircraft equipped with technically enhanced cockpits [T], also known as glass cockpits, which included a large primary flight display and multifunction display with GPS capability. The remaining eleven flights, those with Tomahawk, Cessna 172, and Alpha 160 aircraft, were traditional analogue [A] cockpit displays. The lesson type was fairly consistent amongst the participants, including circuit training, simulated forced landings, and stalling practice. Most participants had completed all of the PPL theory papers, with the English language test being the most common item not yet completed. None of the participants had consumed any significant medication, with only two participants having consumed a Paracetemol-based painkiller earlier in the day that testing took place. Caffeine consumption [Yes/No] was mixed, with six participants from the main group and three control participants consuming caffeine, mostly a single cup of coffee or an energy drink, between one and two hours before the flight. All of the participants except one had eaten within three hours of the flight, an average of an hour and a half previously. The one participant who had not eaten was completing a morning flight.

The five participants in the control group had comparatively very similar ages; the average age was 19 years old, with a range from 18 to 20, and standard deviation of 0.89 years. The flight time of the control group participants was somewhat higher at 92 hours. This is not seen as relevant though, as the control group did not actually complete a training flight. One participant in the control group recorded a low result for recent sleep time, at only four hours, due to a part time job they worked after the previous day's flying.

#	Gender	Age	TFT	RSleep	USleep	Aircraft	Cockpit	Caffeine
1	M	24	45	8	8	C172	A	N
2	F	18	100	7	8	P161	T	N
3	M	23	70	7	8	P112	A	Y
4	M	26	121	8	7	P161	T	N
5	M	18	56	9	8	P181	T	Y
6	M	21	100	8	8	P161	T	N
7	M	24	50	6	6	A160	A	Y
8	M	23	50	8	7	A160	A	Y
9	M	22	136	6	8	C172	A	N
10	M	19	65	6	8	P161	T	Y
11	F	36	95	6	7	P112	A	N
12	M	23	45	7	7	A160	A	N
13	M	24	52	8	8	C172	A	N
14	M	24	52	6	8	C172	A	N
15	M	19	50	8	8	A160	A	N
16	M	18	22	7	8	A160	A	Y
17	M	18	80	4	8	-	-	N
18	M	20	155	7	8	-	-	Y
19	M	19	160	7	8	-	-	Y
20	M	18	42	7	7	-	-	Y
21	M	18	23	7	8	-	-	N

Table 2: Demographic overview of participants

4.2.2. Tools

The questionnaire, having been successfully validated as outlined in chapter three, serves as the subjective method of measuring fatigue. The comparison between pre-flight and post-flight scores on the ten core items will form the main measure of fatigue in the questionnaire. The five items assigned to the pre-flight section are detailed below, in addition to whether the score needs to be reversed prior to calculating an average result:

Pre-flight core item text	Score reversed?
I feel energetic	No
I feel more forgetful than normal	Yes
I can concentrate very well	No
I have an overall feeling of tiredness	Yes
I feel mentally exhausted	Yes

Table 3: Pre-flight core items

Post-flight items are outlined below, with an indication where score reversal is required:

Post-flight core item text	Score reversed?
Mentally, I feel in good condition	No
I find it difficult to concentrate	Yes
I can think clearly	No
I feel relaxed	No
I feel worn out	Yes

Table 4: Post-flight core items

The remaining questionnaire items will serve to further analyse the participant's judgement of the flight [see appendix A].

The PPVT, operated from a laptop computer, is a more specific and objective method of measuring fatigue. The validation study confirmed it as a reliable method which is intuitive and easy for participants to use. While participants complete PPVT for the same amount of time, the number of separate responses can vary, as the time between individual PPVT stimuli also varies, up to a maximum of twelve seconds. While the number of responses varies between participants, it still represents the same amount of time spent undergoing testing. Because of this, the measure of fatigue from the PPVT results will be the difference in average response times between pre-flight PPVT [PPVTpre] and post-flight PPVT [PPVTpost].

4.2.3. Procedure

The ATO which was selected to participate has a detailed online schedule which was used to determine when participants would be flying, and trips to the premises were made at appropriate times in order to test the participants. Usual flight times were four times a day, at 08:00, 10:00, 13:00, and 15:00. In certain situations it was possible to stagger two participants who were flying at the same time, but testing one participant per time slot was by far the more common situation. Due to the quick turnaround and OTP considerations between the 08:00 and 10:00 flights and the 13:00 and 15:00 flights, only one testing session was possible per pair of times. For example, a student departing at 08:00 would not be able to complete the testing with enough time for a student departing at 10:00 to also participate. This effectively limited testing to two students per day. Further complications of inclement weather, scheduling changes and conflicts, operational factors, and occasions where students did not have enough time to complete the testing were fairly common, and all of these considerations meant that the testing was a somewhat lengthy affair, spread over several weeks.

The pre-flight questionnaire section and PPVT were administered as close as possible to the start of the training flight, in an attempt to limit the effect of extraneous variables and restrict the measurement of fatigue to only that which related directly to the flight. After meeting with each participant before their flight, a small briefing room was used to conduct the testing, which ensured the participants were not distracted while testing, and especially the PPVT, was being completed. After the participant completed the questionnaire and was instructed on the use of the PPVT they were left alone in the briefing room for the five minutes it took them to complete the PPVT. After five minutes, as measured by a stopwatch, the participant was advised that testing had been completed and they were free to go. The PEBL software also provided time information with each PPVT response, which allowed further confirmation of the elapsed time. This entire procedure took between ten and fifteen minutes.

A similar procedure was completed after the participant's flight, to administer the post-flight questionnaire section and second PPVT. This completed the testing procedure.

4.2.4 Analysis

IBM SPSS Statistics software [version 21.0] was used to perform the analysis of the data gathered from the research. The primary techniques used for analysing the data were simple descriptive statistics: Correlation coefficient [Pearson's *r*], Confidence intervals [CI], Effect size [Cohen's *d*] (Cohen, 1988), Mean, Standard deviation [SD], and Student's t-test. All t-tests are paired-samples and two-tailed.

4.2.4.1. Questionnaire reliability

The first step in the initial analysis process was to re-assess the reliability of the core items in the questionnaire fatigue scale [QFS], to determine whether the results of a similar analysis done at the pilot study stage are replicated. The reliability analysis was carried out on all ten items discussed in section 4.2. and on the entire sample, including both pre-flight and post-flight testing from both the main and control groups. Cronbach's alpha was calculated at 0.733. This result is only slightly less than the 0.799 score obtained from the pilot test. Table 5 shows a breakdown of the statistics for each of the ten items, including item correlation with the whole scale and alpha coefficient if the item was deleted. Every Likert questionnaire item is allocated a reference code, which identifies whether it is in the pre-flight [PRE] or post-flight [POST] questionnaire section, and what number item it is [See appendix B]

Cronbach's Alpha [α] was used as the means of determining reliability by means of internal consistency amongst the core items. Prior to this analysis, the scores from items PRE04, PRE06, PRE07, POST04, and POST07 were reversed as outlined previously in 4.2. Cronbach's Alpha for all ten core items combined was calculated at 0.733 based on results from all 21 participants in both the main and control groups. This is only slightly less than the 0.799 score obtained in the pilot test. Further analysis was undertaken to determine the inter-correlation between individual items [See Table 5].

The analysis showed that most of the items correlated above 0.3, with the exception of item POST06 which scored only 0.086; if this item was excluded, then alpha would increase to 0.762. The exclusion of one other item, PRE06, would result in a marginal increase of alpha to 0.737. However, the removal of either item does not seem to improve Alpha considerably. Furthermore, inter-reliability analysis of the QFS subscales [PRE and POST] shows that item POST06 actually correlates to the other four POST items quite well [r = 0.41], and its respective alpha would drop by 0.01 should it be removed. This suggests that POST06 does measure a similar construct to the remaining items. Therefore, it was decided that all ten items are still measuring the same construct [subjective fatigue] and that all PRE and POST items can be taken as reliable when assessing changes in self-reported levels of fatigue before and after flying. For each participant, their average score of items PRE03-PRE07 would give the Questionnaire Fatigue Score [pre-flight], or 'QFSpre', and their average score of items POST03-POST07 would give the Questionnaire Fatigue Score [post-flight], or 'QFSpost'.

Item	**Mean**	**SD**	***r***	**α if deleted**
PRE03	5.14	.964	.353	.718
PRE04	5.95	.865	.380	.717
PRE05	5.38	.865	.633	.691
PRE06	4.48	1.834	.307	.737
PRE07	5.38	1.244	.331	.721
POST03	4.86	1.236	.519	.693
POST04	4.48	1.470	.360	.718
POST05	4.71	1.146	.636	.678
POST06	4.95	1.1456	.086	.762
POST07	3.67	1.592	.610	.670

Table 5: Inter-reliability analysis of QFS

4.2.4.2. Variable distribution

Table 6 summarises the relevant descriptive statistics for the research variables, which will be used for 'cleaning up' the data, a method outlined by Tabachnick and Fidell (2001). Items PRE03-PRE07 and POST03-POST07 have been replaced with QFSpre and QFSpost respectively. The main purpose is to determine whether the variables have any missing or incorrect values and whether they are normally distributed or not.

Item	N	Min	Max	Mean	SD	z-skew	z-kurt
PPVTpre	21	285.53	386.40	344.81	24.28	-0.59	0.52
PPVTpost	21	269.00	405.91	341.86	36.91	1.70	-0.51
QFSpre	21	3.60	6.60	5.27	0.82	-1.01	-0.56
QFSpost	21	3.00	6.20	4.53	0.97	-0.16	-1.06
PRE01	21	4.00	7.00	6.06	0.85	-1.53	0.81
PRE02	21	1.00	7.00	4.50	1.71	-1.29	-0.29
PRE08	21	2.00	7.00	5.13	1.46	-0.96	-0.14
POST01	16	3.00	7.00	5.57	1.21	-1.87	0.22
POST02	16	4.00	7.00	5.57	0.98	-0.79	-0.75
POST08	16	1.00	7.00	4.81	1.60	-1.74	0.75
POST09	16	3.00	7.00	5.31	1.14	-1.80	0.41
POST10	16	1.00	5.00	2.81	1.22	-1.68	-0.77
POST11	16	1.00	6.00	2.44	1.67	2.21	0.74
POST12	16	1.00	7.00	3.88	1.54	-0.24	-0.03
POST13	16	1.00	6.00	2.88	1.26	1.28	1.24
POST14	16	1.00	6.00	3.06	1.48	1.28	-0.66
POST15	16	1.00	6.00	2.88	1.63	1.16	-0.81
POST16	16	1.00	7.00	3.19	1.56	2.15	1.37
POST17	16	2.00	7.00	4.50	1.32	0.18	-0.24
POST18	16	2.00	7.00	5.13	1.31	-1.54	1.08
POST19	16	1.00	6.00	4.38	1.31	-1.79	1.52
POST20	16	2.00	7.00	5.38	1.36	-1.40	1.01

Table 6: Initial distribution analysis of results

The number of participants for each particular item, as shown in the 'N' column, is constant among similar items. Those that were completed by both the main and control groups show N = 21, while the remaining items, which only the main group completed, shows N = 16.

Z-scores were calculated for both skewness [z-skew] and kurtosis [z-kurt], following a procedure outlined by Tabachnick and Fidell (2001). A conservative probability of 1% [$p > 0.01$, two-tailed] is used for ascertaining a significant departure from normality, a threshold which is achieved when z-scores for skewness and kurtosis are more extreme than 2.58 or -2.58. As can be seen, the z-scores for skewness and kurtosis [calculated using formulae provided by Tabachnick and Fiddell, 2001], are smaller than the above thresholds and can therefore be considered as normally distributed. This relative normality of the variables supports the use of a parametric approach for further analysis of the data.

4.3. Results

The results are split into two main sections. The first section addresses the main variables, subjective fatigue as measured by the variable pair of QFSpre and QFSpost, and objective fatigue as measured by PPVTpre and PPVTpost. The second section covers the remaining questionnaire items.

4.3.1. Main measures of fatigue

As previously mentioned, the two main measures of fatigue in this study are QFS and PPVT, which will form the basis of assessing fatigue amongst the participants. Figures and tables for control group data are smaller and centred to allow for easier identification.

4.3.1.1. Subjective fatigue [QFS]

Subjective fatigue, as measured by QFSpre and QFSpost, increased for the main group, with QFS dropping from an average of 5.58 to an average of 4.4 [on a seven-point Likert scale with 7 being the least fatigued, or most alert]. The results show a considerable drop in the fatigue score between pre-flight and post-flight testing, indicating an increase in fatigue. There is also a strong correlation between the two [$r = 0.64$, $p = 0.008$], and the results are both highly practically and statistically significant [$d = 1.93$, $t(15) = 6.315$, $p < 0.001$].

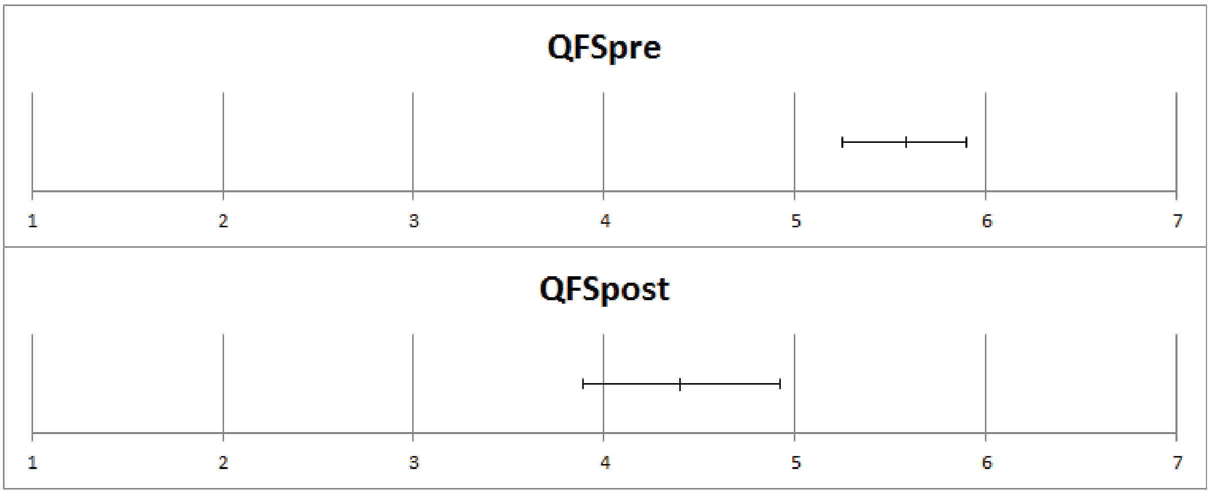

Figure 1: Main group QFS scores

[Main]	Mean	CI	SD
QFSpre	5.58	[5.25 - 5.90]	0.61
QFSpost	4.40	[3.89 – 4.92]	0.97
Association	$r = 0.64$, $p = 0.008$		
Effect size	$d = 1.93$		
Significance	$t(15) = 6.315$, $p < 0.001$		

Table 7: Main group QFS results

By comparison, results from the control group show an increasing score from 4.28 to 4.96, indicating a decreasing level of fatigue. However, due to the small number of participants in the control group, the confidence intervals are somewhat larger. Furthermore, correlation [$r = 0.40$, $p = 0.499$], and practical and statistical significance [$d = 1.08$, $t(4) = -1.667$, $p = 0.171$] are all considerably lower than the main group.

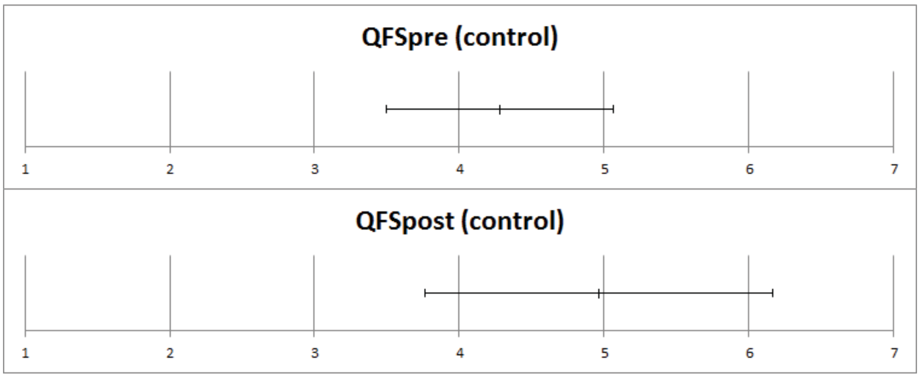

Figure 2: Control group QFS scores

[Control]	**Mean**	**CI**	**SD**
QFSpre	4.28	[3.50 – 5.06]	0.63
QFSpost	4.96	[3.76 – 6.16]	0.96
Association	*r* = 0.40, *p* = 0.499		
Effect size	*d* = 1.08		
Significance	*t*(4) = -1.667, *p* = 0.171		

Table 8: Control group QFS results

The above results clearly show a contrast between the main and control groups. For figure 3, the negative gradient of the main group shows a decrease in QFS and therefore an increase in fatigue, while the positive gradient of the control group shows an increase in QFS and therefore a decrease in fatigue. In order to more fully understand the scale of the difference between the two groups, a differential effect size can be calculated by considering all four results, taking into account the initial fatigue levels of participants from both groups. The process is similar to that which was used to calculate individual effect sizes for the main and control groups separately, except all four results are considered, that is QFSpre and QFSpost for both main and control groups. As the means for each group change in opposite directions, the differential effect size is consequently very high, *d* = 2.95.

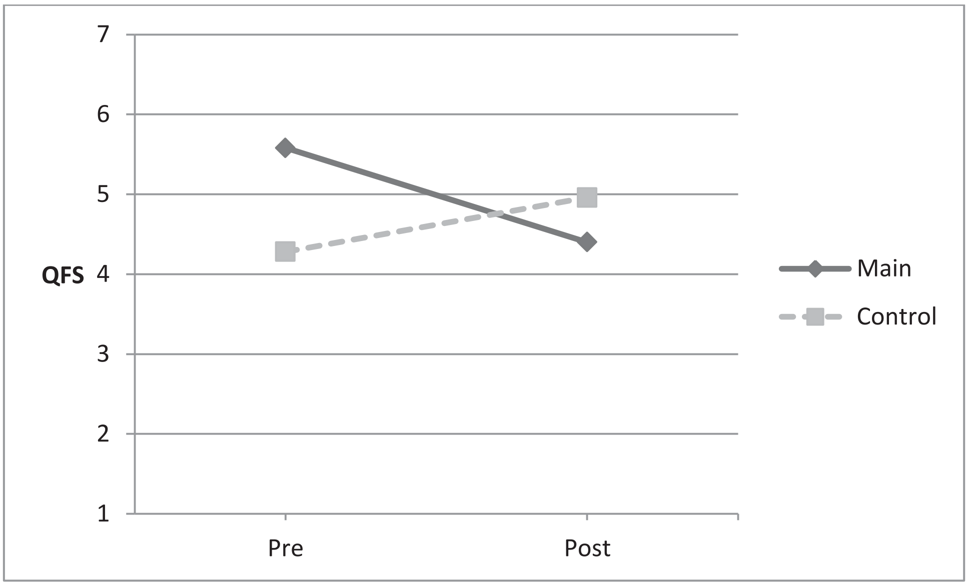

Figure 3: Main and control group QFS changes

[QFS]	Pre	Post
Control	4.28	4.96
Main	5.58	4.40

Table 9: Summary of QFS scores

4.3.1.2. Objective fatigue [PPVT]

As mentioned above, PPVT data will be primarily analysed with respect to the average response times from each five-minute pre-flight [PPVTpre] and post-flight [PPVTpost] testing period. Average scores with 95% confidence intervals are displayed in the figure below. The scale has been oriented for consistency with the QFS score figures as displayed above; a worsening reaction time, moving to the left of the scale, indicates higher fatigue.

With an increase of only 3ms, there is no significant difference between PPVTpre and PPVTpost for the main group. While correlation is high [r = 0.75, p = 0.001], practical and statistical significance results [d = 0.06, $t(15)$ = -0.675, p = 0.510] are very poor.

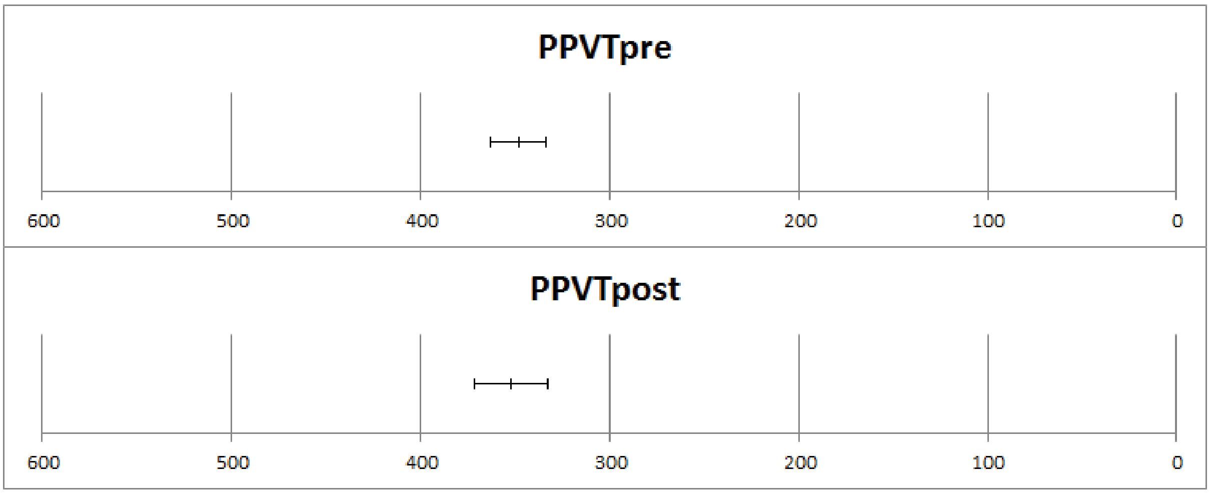

Figure 4: Main group PPVT scores

[Main]	Mean	CI	SD
PPVTpre	348ms	[334 – 362]	26.44
PPVTpost	352ms	[332 – 371]	36.46
Association	r = 0.75, p = 0.001		
Effect size	d = 0.06		
Significance	$t(15)$ = -0.675, p = 0.510		

Table 10: Main group PPVT results

The results of the PPVT from participants in the control group differ from the main group, a large decrease in average score, by 26ms, indicating that fatigue levels actually decreased during the intervening period. The extremely high correlation [r = 0.91, p = 0.03], a medium effect size [d = 0.39], and statistically significant results [$t(4)$ = 10.085, p = 0.001] are promising.

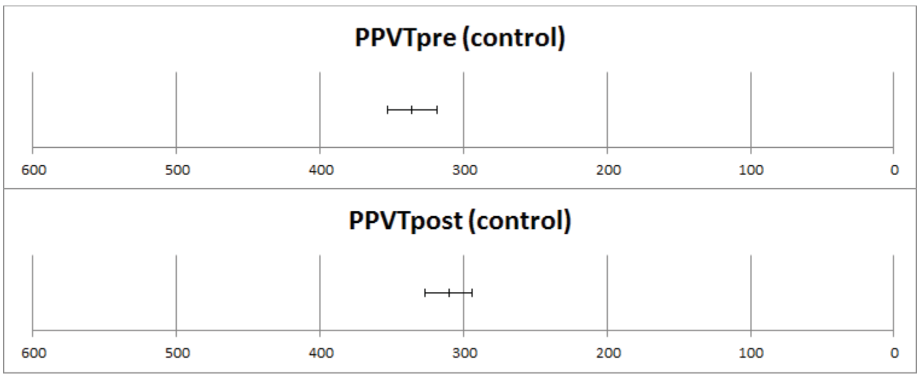

Figure 5: Control group PPVT scores

[Control]	**Mean**	**CI**	**SD**
PPVTpre	336ms	[319 – 353]	13.71
PPVTpost	310ms	[294 – 327]	13.29
Association	$r = 0.91, p = 0.03$		
Effect size	$d = 0.39$		
Significance	$t(4) = 10.085, p = 0.001$		

Table 11: Control group PPVT results

Figure 6 has been oriented with a reversed y-axis to maintain consistency with the similar figure 3 for QFS results above. The slight negative gradient of the main group's results indicates a worsening reaction time, and hence higher PPVT values and higher fatigue, while a positive gradient, as shown in the control group results, indicates the opposite. There was a low effect size result for the main group and a moderate result from the control group, d = 0.06 and d = 0.39 respectively. However, seeing that the control group improved while the main group remained mostly constant, the differential effect size between the two groups is much larger, at d = 2.15.

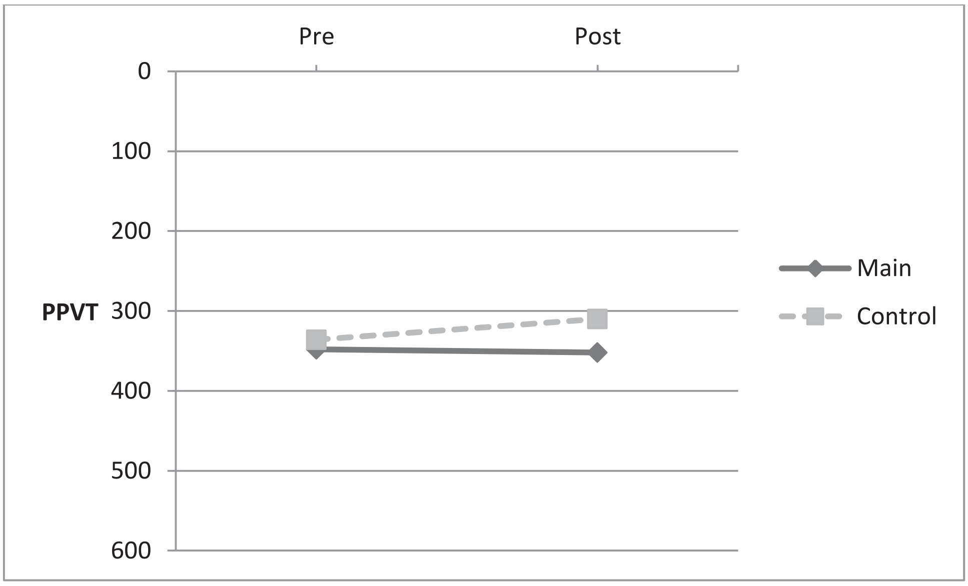

Figure 6: Main and control group PPVT changes

[PPVT]	Pre	Post
Control	336ms	310ms
Main	348ms	352ms

Table 12: Summary of PPVT scores

4.3.1.3. Correlations between QFS and PPVT

There are medium to strong correlations between pre and post QFS and PPVT measurements for both main and control groups [r = 0.31 to 0.52], but significance levels are low [p = 0.084 to 0.544] due to the small sample sizes.

[Main]	*r*	*p*
QFSpre & PPVTpre	0.45	0.084
QFSpost & PPVTpost	0.31	0.238

Table 13: Main group correlations: QFS vs PPVT

[Control]	*r*	*p*
QFSpre & PPVTpre	0.37	0.544
QFSpost & PPVTpost	0.52	0.372

Table 14: Control group correlations: QFS vs PPVT

4.3.1.4. Correlations between main variables and demographics

In order to calculate correlations between the main variables and other factors it is necessary to construct two additional variables, QFSdiff [QFSpre – QFSpost] and PPVTdiff [PPVTpre – PPVTpost]. This represents the change in fatigue as represented by the change in main variable scores between pre-flight and post-flight testing for both the main and control groups. These variables are constructed so that a positive correlation with a second variable indicates an increase in fatigue as the second variable increases.

Some correlations between the main variables and demographics are not relevant for the control group, because no training flight was completed, and in the case of gender because all of the participants in the control group were male [See Table 16]. While there are several strong correlations between the main variables [QFS and PPVT] and demographic data, few are significant given the small sample sizes [See

Tables 15 and 16]. In fact, only three correlations are statistically significant, and even then only at the 5% level. The first, and the only in the main group, age vs QFS from the main group, is negatively correlated [r = -0.51], which suggests that younger participants recorded more of a drop in QFS between pre-flight and post-flight testing. There was also a high correlation between age and QFS in the control group [r = -0.71], but the very small sample size returned a low significance [p = 0.179], albeit one of the highest in the group.

Total flight time [TFT] and QFS returned a significant negative correlation in the control group [r = -0.91, p = 0.032], which could suggest that participants with less flight time, and hence at earlier stages in their training, are more likely to succumb to fatigue. However, it is questionable whether TFT is even relevant to the control group since they did not actually complete a training flight. However, the main group did also return a high correlation [r = -0.48], but the p level [0.059] was slightly too high for the result to be considered significant.

Finally, recent sleep and PPVT were strongly positively correlated [r = 0.93, p = 0.022], which can likely be explained by examining individual results. One participant had a very small amount of recent sleep, only four hours, and recorded a large improvement in their PPVT scores, while the other four participants had fairly consistent and average results for both variables.

[Main]	QFSdiff		PPVTdiff	
	r	*p*	*r*	*p*
Age	-0.51	0.044	-0.32	0.227
Gender	-0.04	0.988	-0.34	0.198
TFT	-0.48	0.059	-0.43	0.096
Recent sleep	-0.23	0.391	0.25	0.350
Usual sleep	0.12	0.658	0.23	0.391
Aircraft type	0.34	0.198	0.00	1.000
Cockpit type	0.02	0.941	0.04	0.883
Theory papers	0.29	0.276	-0.12	0.658
Caffeine	0.39	0.135	0.17	0.529
Food	-0.15	0.579	0.23	0.391

Table 15: Main group correlations: QFS and PPVT vs Demographics

[Control]	QFSdiff		PPVTdiff	
	r	*p*	*r*	*p*
Age	-0.71	0.179	0.20	0.747
Gender	---	---	---	---
TFT	-0.91	0.032	-0.04	0.949
Recent sleep	0.24	0.697	0.93	0.022
Usual sleep	-0.71	0.179	-0.05	0.936
Aircraft type	---	---	---	---
Cockpit type	---	---	---	---
Theory papers	---	---	---	---
Caffeine	-0.37	0.540	0.33	0.588
Food	-0.78	0.120	0.28	0.648

Table 16: Control group correlations: QFS and PPVT vs Demographics

4.3.2. Auxiliary questionnaire items

The remaining auxiliary questionnaire items [AQI] will be assessed in a similar way as the main research variables. It is important to note that none of these variables have had their scores reversed as was required for the construction of the QFS variables. As a result, they cannot be interpreted in a manner whereby a lower score indicates more fatigue. Furthermore, all variables mentioned in this section are only from the main group, as the control group could not complete them due to the lack of an intervening training flight. The items will be analysed in relation to six broad topics: training flight content, workload, preparation level, future flights, general performance measures, and specific performance measures.

4.3.2.1. Training flight content

Two items, posed to participants at the pre-flight stage, were in relation to the content of their training flight: **"There are new concepts or techniques featuring in the flight" [PRE02]**, and **"The flight will mostly be revision of concepts and techniques that I am already familiar with" [PRE08].**

Some participants responded positively to both items, indicating that the flight built on previous knowledge in addition to including new material. Others responded negatively to the first item and positively to the second, indicating that only revision of previously learnt techniques was to feature in the training flight. Finally, some responses were the opposite, indicating that they were learning new material in their flight.

Overall, the main group of participants were somewhat agreeable with the statement that they will be learning something new, and more agreeable with the statement that they were revising concepts and techniques already learnt.

	Score	**CI**	**SD**
PRE02	4.5	[3.59 – 5.41]	1.71
PRE08	5.13	[4.35 – 5.90]	1.45

Table 17: Overview of PRE02 and PRE08

4.3.2.2. Workload

The item "**There was a high workload throughout the flight" [POST08],** relates to the workload throughout the training flight. Responses were overwhelmingly positive; only one participant gave a negative response to this item. A high workload was predicted in the literature review and as such there is a strong possibility that participants suffered from some degree of cognitive fatigue.

	Score	CI	SD
POST08	5.13	[4.48 - 5.77]	1.20

Table 18: Overview of POST08

4.3.2.3. Preparation level

Items PRE01 "**Mentally I am well prepared to undergo the flight"** and POST02 **"I was well prepared to undergo the flight"** are a pairing which provides interesting results. Comparing them can give an indication as to how the participants viewed their level of preparation both before and after the flight. Responses to the first item indicated that the participants thought they were well prepared mentally for the training flight. All participants except one gave a positive response, with the one other participant giving an 'average' score of 4.

The second item showed a drop of 0.5 compared to the similar item PRE01 score of 6.06. Standard deviations were about the same for both items. This indicates that some participants may have possibly over-estimated their level of preparation in responding to the first question, and were more accurate in judging this level after the flight

	Score	CI	SD
PRE01	6.06	[5.61 – 6.52]	0.85
POST02	5.56	[5.09 – 6.04]	0.89

Table 19: Overview of PRE01 and POST02

4.3.2.4. Future flights

Three more items build on the assessment of participant preparation level to evaluate a proposed future training flight: **"I am pleased with the outcome of the flight" [POST01], "I am well prepared to complete another training flight commencing now" [POST19],** and **"I would need to have some time to rest and recover before completing another similar flight today" [POST20].**

The first item returns a very high score of 5.63, and the overall affirmative response indicates that the participants did not experience any disappointments throughout the flight, and are therefore unlikely to be affected by emotional fatigue, as a result of, for example, failing to grasp important concepts or techniques. It means that emotional fatigue can, to a certain degree, be ruled out as a significant extraneous variable, and therefore any fatigue is more likely to be caused by cognitive factors.

The second item shows a reasonably large drop compared to the participants' evaluation of their preparedness of the initial flight as judged before the flight [PRE01], a drop of 1.62 from 6.06 to 4.38, and after the flight [POST02], a drop of 1.12 from 5.56 to 4.38. This is reinforced by the overall affirmative response to the third item, which indicates a need to "rest and recover" before attempting another training flight.

	Score	**CI**	**SD**
POST01	5.63	[4.93 – 6.32]	1.31
POST19	4.38	[3.68 – 5.07]	1.31
POST20	5.38	[4.65 – 6.10]	1.36

Table 20: Overview of POST01, POST19, and POST20

4.3.2.5. General performance measures

Four items refer to performance as a general concept: **"My performance gradually decreased throughout the flight" [POST10], "My performance sharply decreased near the end of the flight" [POST11], "I would be able to perform at the same level of proficiency on another similar training flight to be commenced now" [POST12],** and **"I would be able to perform at a higher level on another similar training flight to be commenced now" [POST13].**

Starting with the third item, this relates to the participants' judgement of their ability to perform on another training flight to be commenced immediately. The response to this item was overwhelmingly negative, but the nature of the item is somewhat ambiguous, as a negative response could relate to performing either better or worse. However, the other data are indicative that any change would be towards a lower level of proficiency; the fourth item specifically suggests this is so. The overall highly negative response to this item, of 2.88, further suggests that a higher level of performance on an immediate subsequent flight would not be likely. Only one participant suggested they would be able to perform at a higher level. Should negative responses to the third item actually relate to a higher level of performance, then a negative correlation would be expected between these two items, however analysis did not show any such correlation whatsoever.

Interestingly, the first two items which sought to determine the nature of any performance decrease at different stages of the flight did not seem to produce any results. Despite the participants showing a considerable drop in their performance levels by comparing performance assessments from before and after the flight as explained above, they did not seem to recognise this as readily when asked to evaluate it by thinking back to the events over the course of the flight.

	Score	**CI**	**SD**
POST10	2.81	[2.16 – 3.46]	1.22
POST11	2.44	[1.55 – 3.33]	1.67
POST12	3.88	[3.05 – 4.70]	1.54
POST13	2.88	[2.20 – 3.55]	1.26

Table 21: Overview of POST10, POST11, POST12, and POST13

4.3.2.6. Specific performance measures

Finally, six items relate to performance as measured by specific factors: **"I found it easy to recall important points and concepts relating to the training flight" [POST09], "I became easily distracted at times near the end of the flight" [POST14], "I made small mistakes or forgot things at the start of the flight" [POST15], "I made small mistakes or forgot things near the end of the flight" [POST16], "My flying ability improved over the course of the flight" [POST17],** and **"I was able to maintain a focus on the requirements of the training flight throughout its duration" [POST18].**

The average results show little difference between the third and fourth items which cover making mistakes or forgetting things, with only an increase of 0.31 between them. Individual results to the items varied, with seven participants indicating a slight increase between the two, four others no change, and the last four a decrease.

With an overall positive response to the first item, only two participants indicated a [slight] negative response, and the final sixth item which is similar, had two negative responses.

The second and fifth items showed varied results.

	Score	CI	SD
POST09	5.31	[4.71 – 5.92]	1.31
POST14	3.06	[2.27 – 3.85]	1.48
POST15	2.88	[2.01 – 3.74]	1.63
POST16	3.19	[2.36 – 4.02]	1.56
POST17	4.50	[3.80 – 5.20]	1.32
POST18	5.13	[4.43 – 5.82]	1.31

Table 22: Overview of POST09, POST14, POST15, POST16, POST17, and POST18

4.3.2.7. AQI Correlations

Out of the 136 total AQI correlations, 62, or 46%, were of medium strength or higher, that is $r > 0.29$ [See Tables 23 and 24]. All of these correlations were in the direction that would be expected from information covered in the literature review, further validating the questionnaire design. For example, participants who found it easy to **"recall important points and concepts relating to the training flight" [POST09]** also indicated that their **"flying ability improved over the course of the flight" [POST17, *r* = 0.56]**, and that they **"were well prepared to undergo the flight" [POST02, *r* = 0.60].** Seventeen correlations of $0.63 > r > 0.49$ are statistically significant at the 5% level [*], while a further 12 correlations of $r > 0.62$ are significant at the 1% level [**]. Some particularly relevant correlations will be evaluated below.

Two strong correlations and one medium correlation regarding **"new concepts or techniques featuring in the flight" [PRE02],** suggest that after completing these types of training flights featuring new content, participants are less likely to be **"prepared to complete another training flight" [POST19, *r* = -0.51],** and to **"perform at a higher level" [POST13, *r* = -0.68],** likely needing **"some time to rest and recover" [POST20 *r* = 0.66]** first.

Participants who found their **"performance gradually decreas[ing] throughout the flight" [POST10]**, also found that it **"sharply decreased near the end" [POST11, *r* = 0.73]**, and that they "**became easily distracted at times near the end of the flight**" **[POST14, *r* = 0.60]**. A **"high workload throughout the flight" [POST08]** was likely to be associated with **"[making] small mistakes or forg[etting] things near the end of the flight" [POST16, *r* = 0.59],** and this was likely to be related to not being able to complete another training flight **"at the same level of proficiency" [POST12, *r* = 0.60],** or higher.

Being **"well prepared to undergo the flight" [POST02]** was linked with being able to easily **"recall important points and concepts relating to the training flight" [POST09, *r* = 0.60],** and with the participant being **"pleased with the outcome of the flight" [POST01, *r* = 0.54]** and being **"well prepared to complete another training flight" [POST19, *r* = 0.72].**

	PRE01	PRE02	PRE08	POST01	POST02	POST08	POST09	POST10
PRE01								
PRE02	0.30							
PRE08	-0.22	-0.75**						
POST01	0.26	0.27	-0.36					
POST02	0.30	-0.24	-0.11	0.54*				
POST08	-0.20	0.13	-0.28	0.20	-0.38			
POST09	0.05	0.09	0.02	0.67**	0.60*	-0.23		
POST10	-0.18	0.08	0.05	-0.42	-0.20	-0.21	-0.15	
POST11	-0.02	0.15	0.11	-0.59*	-0.49	-0.06	-0.36	0.73**
POST12	-0.04	-0.56*	0.27	-0.06	0.49	-0.42	0.10	-0.54*
POST13	-0.43	-0.68**	0.41	-0.23	0.13	-0.12	-0.06	-0.19
POST14	-0.16	0.18	0.12	0.05	-0.28	0.15	0.15	0.60*
POST15	-0.52*	-0.19	-0.02	-0.56*	-0.45	0.21	-0.70**	0.09
POST16	-0.01	0.16	-0.28	-0.19	-0.32	0.59*	-0.41	0.44
POST17	-0.27	-0.47	0.52*	0.19	0.37	-0.34	0.56*	-0.15
POST18	0.29	0.18	-0.01	0.30	0.34	-0.39	0.37	-0.23
POST19	-0.08	-0.51*	0.11	0.09	0.72**	-0.33	0.36	-0.08
POST20	0.15	0.66**	-0.36	0.01	-0.41	0.17	-0.08	0.29

Table 23: Correlation matrix for AQI, part 1: PRE01 to POST10

	POST11	POST12	POST13	POST14	POST15	POST16	POST17	POST18	POST19
POST12	-0.55*								
POST13	-0.16	0.71**							
POST14	0.55*	-0.64**	-0.14						
POST15	0.12	0.15	0.45	-0.19					
POST16	0.61*	-0.60*	-0.19	0.43	0.19				
POST17	-0.29	0.36	0.32	0.09	-0.19	-0.44			
POST18	-0.09	0.04	-0.31	-0.11	-0.34	-0.34	0.35		
POST19	-0.20	0.62*	0.56*	-0.32	-0.07	-0.23	0.23	0.05	
POST20	0.28	-0.77**	-0.87**	0.22	-0.25	0.25	-0.3	0.20	-0.65**

Table 24: Correlation matrix for AQI, part 2: POST11 to POST19

4.3.2.8. Correlations between main variables and AQI

Out of the 36 total correlations, 13, or 36%, were of medium strength or higher, that is $r > 0.29$ [See Table 25]. All of the correlations with QFS were in the direction that would be expected from information covered in the literature review and previous analysis of results, while PPVT results did not provide any significant results. Five correlations of $0.63 > r > 0.49$ are statistically significant at the 5% level, while a further two correlations of $r > 0.62$ are significant at the 1% level, all of these being with QFS. These seven correlations will be evaluated below. Correlations are between AQI and an increasing level of fatigue as measured by either QFS or PPVT. For example, the weakly positive correlation between PRE02 and QFS indicates that affirmative responses to PRE02 are positively correlated with a higher level of fatigue as measured by QFS.

Participants who recorded a smaller difference between their QFS scores, indicating a lower level of fatigue, were likely to have been **"well prepared to undergo the flight" [POST02, *r* = -0.50]**, and **"found it easy to recall important points and concepts relating to the training flight" [POST09, *r* = -0.57].** They were also likely to be **"well prepared to complete another training flight" [POST19, *r* = -0.71]**, **"at the same level of proficiency" [POST12, *r* = -0.64]**, or **"higher" [POST13, *r* = -0.53].** However, participants who recorded a higher level of fatigue were more likely to **"[make] small mistakes or forget things near the end of the flight" [POST16, *r* = 0.58],** and likely need **"some time to rest and recover before completing another similar training flight" [POST20, *r* = 0.55].**

Finally, while only just not statistically significant by a margin of $p = 0.02$, POST08 and QFS provide an interesting correlation, whereby a **"high workload throughout the flight"** is more likely to be associated with a high QFS.

	QFSdiff		**PPVTdiff**	
	r	***p***	***r***	***p***
PRE01	0.23	0.391	0.42	0.105
PRE02	0.20	0.458	0.19	0.481
PRE08	-0.12	0.658	-0.18	0.505
POST01	-0.12	0.658	0.32	0.227
POST02	-0.50*	0.049	-0.10	0.713
POST08	0.48	0.060	0.21	0.435
POST09	-0.57*	0.021	-0.19	0.481
POST10	0.16	0.554	-0.35	0.184
POST11	0.32	0.227	-0.09	0.740
POST12	-0.64**	0.008	-0.22	0.413
POST13	-0.53*	0.035	-0.31	0.243
POST14	0.21	0.435	-0.11	0.685
POST15	0.16	0.554	-0.16	0.554
POST16	0.58*	0.019	0.05	0.854
POST17	-0.29	0.276	-0.12	0.658
POST18	-0.05	0.854	0.24	0.371
POST19	-0.71**	0.002	-0.26	0.331
POST20	0.55*	0.027	0.36	0.171

Table 25: Correlations: QFS and PPVT vs AQI

4.3.2.9. Correlations between AQI and demographics

Out of the 180 total correlations between AQI and demographics, eleven of $0.63 > r > 0.49$ are statistically significant at the 5% level, while a further six correlations of $r > 0.62$ are significant at the 1% level [See Tables 26 and 27]. Some particularly relevant correlations will be evaluated below.

Participants with more total flight time **[TFT]** were more likely to be completing a training flight featuring a **"revision of [familiar] concepts or techniques" [PRE08, *r* = 0.52]**, and were also more likely to find it **"easy to recall**

important points or concepts relating to the training flight" [POST09, *r* = 0.52]. Higher TFT also correlated with an improving **"flying ability … over the course of the flight" [POST17, *r* = 0.64].**

There were several strong correlations with recent sleep in the past 24 hours [RecentS], and usual amount of sleep each night [UsualS]. More sleep the previous night [RecentS] was correlated with a feeling of being **"well prepared to undergo the flight" [POST02, *r* = 0.56]**, and being **"well prepared to complete another training flight" [POST19, *r* = 0.62].** Less sleep the previous night was likely to be associated with **"performance sharply decreas[ing] near the end of the flight" [POST11, *r* = -0.50],** and also becoming **"easily distracted at times near the end of the flight" [POST14, *r* = -0.51]**. Participants who usually slept for more hours each night were also more likely to feel **"well prepared to undergo the flight" [PRE01, *r* = 0.68],** and less likely to make **"small mistakes or [forget] things near the start of the flight" [POST15, *r* = -0.51].**

Finally, increased caffeine consumption was moderately correlated with having difficulty recalling **"important points and concepts relating to the training flight" [POST09, *r* = -0.54]**.

There were no notable correlations between the remaining demographic data.

	Age	Gender	TFT	Recent sleep	Usual sleep
PRE01	-0.56*	-0.49	-0.2	0.07	0.68**
PRE02	-0.12	0	-0.28	-0.18	0.44
PRE08	0.29	0.24	0.52*	-0.34	-0.39
POST01	0.17	-0.04	0.02	0.32	0.39
POST02	0.11	-0.25	0.15	0.56*	0.29
POST08	-0.19	-0.04	-0.44	0.04	-0.02
POST09	0.46	0.24	0.52*	0.12	0.37
POST10	-0.03	0.38	0.26	-0.30	0.08
POST11	-0.10	0.02	0.10	-0.50*	0.04
POST12	0.10	-0.35	0.19	0.41	-0.33
POST13	0.10	-0.12	0.44	0.29	-0.49
POST14	0.21	0.38	0.40	-0.51*	-0.05
POST15	-0.21	0.03	-0.15	0.18	-0.51*
POST16	-0.39	-0.17	-0.30	-0.16	0.15
POST17	0.37	0.30	0.64**	0.08	-0.16
POST18	0.29	0.11	-0.02	0.03	0.23
POST19	0.19	-0.26	0.23	0.62*	0.02
POST20	-0.09	0.32	-0.42	-0.31	0.34

Table 26: Correlation matrix part 1: Demographics vs AQI

	Cockpit type	Theory papers	Caffeine	Food
PRE01	-0.05	-0.41	-0.06	0.02
PRE02	-0.04	-0.47	0.18	0.07
PRE08	0.04	0.62*	-0.27	0.16
POST01	0.20	-0.13	-0.24	0.11
POST02	-0.13	-0.24	-0.48	-0.10
POST08	0.16	0.08	0.49	0.03
POST09	0.05	-0.02	-0.54*	0.16
POST10	-0.12	0.31	-0.01	-0.32
POST11	-0.27	0.21	0.25	0.06
POST12	0.06	-0.07	-0.23	0.01
POST13	0.29	0.45	0.02	-0.02
POST14	-0.03	0.66**	0.17	-0.17
POST15	0.31	0.10	0.47	-0.17
POST16	-0.17	0.10	0.34	-0.06
POST17	0.16	0.30	-0.46	-0.04
POST18	-0.28	-0.43	-0.36	0.49
POST19	-0.09	-0.08	-0.27	0.09
POST20	-0.29	-0.37	0.01	0.08

Table 27: Correlation matrix part 2: Demographics vs AQI

Chapter five – Discussion

5.1. Study goals

The aim of this study was to investigate cognitive fatigue that accumulated over the course of a one-hour training flight, with the ultimate goal of using the data gathered to set a baseline that could be used in future studies. Results from both the subjective and objective tools indicated that fatigue was indeed present among the participants in the main group following the training flight. In particular, QFS returned some excellent results, and is the best candidate for setting a baseline for cognitive fatigue in student pilots.

5.2. Research questions

This section will interpret the results of the main study primarily by addressing the research questions that were established at the beginning of chapter four.

<u>5.2.1. Do student pilots demonstrate any fatigue as measured by a subjective tool?</u>
The questionnaire results were promising, with QFS data from the main group indicating a significant increase in fatigue over the course of the training flight. They also showed that every single participant in the main group scored lower on QFSpost than their QFSpre score, in other words, every participant from the main group was more fatigued after their flight, as measured by QFS. The average QFS decrease was 1.18 points from 5.58 to 4.4, which represents a 20% drop. However in the control group, QFS increased by an average of 0.68 points, from 4.28 to 4.96, or an 11% improvement. To have both a large and consistent difference between pre-flight and post-flight scores, coupled with the contrasting results from the control group, is a strong indication that QFS measured considerable levels of fatigue. These results are very promising for the further use of QFS as a measure of subjective fatigue.

Results to other questionnaire items made by the main group participants also show strong indications of fatigue, for example all of the participants except one agreed with a high workload being present throughout the training flight [POST08, score =

5.14]. Workload has already been discussed as a strong predictor of cognitive fatigue (Gartner & Murphy, 1976; Powell et al., 2007; Yen et al., 2009). In light of the overall affirmative responses to a high workload, it is therefore likely that the fatigue measured by QFS is indeed cognitive fatigue, and not resulting from other sources. Workload was also found to correlate with participants making small mistakes or forgetting things near the end of the flight [$r = 0.59$, $p < 0.05$], another indication of cognitive fatigue (Caldwell et al., 2003; Fletcher, et al., 2003).

Participants were mostly pleased with the outcome of the flight [POST01, score = 5.63], which is an indication that emotional fatigue is not likely to be a factor. If the participants experienced disappointing moments throughout the flight, for example by failing to grasp important concepts, or not being able to master important techniques, then this might reflect on QFSpost items such as "I find it difficult to concentrate", "I can think clearly", and "I feel relaxed". As this is not the case, it can be stated with an even higher level of confidence that any fatigue experienced by the participants is cognitive fatigue.

An important comparison between some of the questionnaire items is the change in level of preparedness among participants. One of the highest scores came from participants strongly agreeing that they were well prepared to undergo the flight [PRE01, score 6.06]. After the flight, when again asked about their level of preparedness to undergo the same flight, the score dropped by 0.5 to 5.56 [item POST02]. This suggests that participants were optimistic about their level of preparation before undergoing the flight, but afterwards when thinking back to how well prepared they actually were, they do not agree to the same extent and consequently produce a more accurate evaluation of their level of preparedness. However, when asked in the post-flight section of the questionnaire whether the participant would be well prepared to complete a second training flight, the level of preparedness falls a further 1.18 points to 4.38 [item POST19]. These figures are almost identical to the change in QFS scores, both having a difference of exactly 1.18 points [See Table 28]. This all but perfect correlation is an important result, and further confirms the design of the questionnaire. It means that a student pilot's preparedness to undergo a training flight at any particular time is strongly associated with their level of fatigue.

	QFS	**Preparedness**
Pre-flight	5.58	5.56
Post-flight	4.4	4.38
Change	1.18	1.18

Table 28: Comparing QFS and preparedness

In post-flight testing, participants agreed that they would need some time to rest and recover first, before undertaking a second training flight as outlined above [POST20, score = 5.38]. This is yet another indication that cognitive fatigue is being measured, as it ties in with the previously established debilitating effects of fatigue.

Several further correlations between different questionnaire items as outlined in section 4.3.2.7. support the nature of the fatigue being measured. When a participant's training flight featured more new concepts or techniques they were less likely to be prepared to go on and complete a second training flight [$r = -0.51$, $p < 0.05$] and to perform at a higher level in doing so [$r = -0.68$, $p < 0.01$]. They were likely to need some time to rest and recover [$r = 0.66$, $p < 0.01$] first. Considering the nature of cognitive fatigue and its relationship with human information processing in a flight training environment, this result is to be expected. The addition of new information, requiring heavy use of working memory, is more likely to place strain on a student's ability to cope with the mental challenges of the training flight and hasten the onset of cognitive fatigue. While results did not show strong levels of acceptance with specific measures of performance impairment, there were nonetheless some strong correlations between these items. A gradually decreasing performance throughout the training flight was closely linked with a sharp decrease in performance [$r = 0.73$, $p < 0.01$] and becoming easily distracted [$r = 0.60$, $p < 0.05$] near the end of the flight. Distraction was also correlated to workload [$r = 0.59$, $p < 0.05$].

Increased fatigue as measured by QFS was likely to result in errors or memory problems [$r = 0.58$, $p < 0.05$], while participants who were not as fatigued as measured by QFS were likely to have found it easy to recall information throughout the flight [$r = -0.57$, $p < 0.05$], have high preparation levels on a second upcoming flight [$r = -0.71$, $p < 0.01$], and to be able to perform at the same level [$r = -0.64$, $p <$

0.01], or higher [$r = -0.53$, $p < 0.01$]. All of these correlations are strong signs of the presence, and lack, of fatigue as appropriate. The findings suggest that cognitive fatigue is indeed the construct being measured, and that there is a very strong indication that the student pilot participants demonstrated fatigue as measured by the questionnaire.

5.2.2. Do student pilots demonstrate any fatigue as measured by an objective tool?

The PPVT gave less convincing results compared to the questionnaire, with the main group only being 3ms slower between PPVTpost and PPVTpre. However, just like as for the questionnaire, the control group showed contrasting results, with the participants actually improving between testing by 25ms.

This difference can be possibly explained by considering the nature of the variable being measured, reaction time. It has been established that fatigue has many effects, and just one of these is impaired reaction time. However, fatigue is not necessarily the only factor that would affect reaction time, and so there is a possibility that it may be affected by an extraneous variable. It is also important to note that it is actually 'measured reaction time' that is being recorded, not actual best possible reaction time of each participant. For example, a participant might not be fatigued, and have a good reaction time, but might not be responding as quickly as they are able to due to distraction or lack of motivation. The participants may be thinking about their upcoming flight, causing distraction and a worse than expected score on the pre-flight test. Comparatively, after having completed the flight they might not be distracted in the same way, and this reduction in distraction serves to counteract some of the impairment caused by fatigue. If this is the case, and pre-flight scores are being artificially impaired, then it would explain why 'post-flight' scores are higher than expected for both the main and control groups.

It is also possible that a degree of 'learning effect' may be taking place, whereby participants are more familiar with the test upon their second attempt and therefore score better than they otherwise would have. This may further explain why the 'post-flight' results for both the main and control groups are higher than expected. The main group shows a significant increase in fatigue as measured by QFS, but only a marginal and insignificant increase in their reaction time. The control group shows a slight decrease in fatigue as measured by QFS, but a large improvement on the reaction time test. The combination of a learning effect and extraneous variables could be skewing PPVT results. Comparing the change between pre-flight and post-flight results in the main and control groups still shows a significant difference between results.

While it could not be explicitly shown that participants were affected by fatigue based on PPVT data, the results do appear to indicate that the main group performed worse than the control group. When also considering the possibility of extraneous variables, this does seem to indicate the presence of fatigue, although arguably not as directly as the QFS data.

5.2.3. Do the results of the subjective and objective tools correlate?

Both QFS and PPVT showed a considerable difference between pre-flight and post-flight testing for the main and control groups. As previously mentioned, it was not possible to find any significant correlation between both sets of data. This is likely due to other factors such as extraneous variables or learning effects skewing the PPVT data such that it does not match with the questionnaire results.

5.2.4. Are the results practically significant?

Both main group and control group QFS results returned high effect sizes, d = 1.93 and d = 1.08 respectively. Differential effect size between main and control groups was even larger, at d = 2.95. Conversely, PPVT data had much poorer effect size results, with only d = 0.06 for the main group and d = 0.39 for the control group. The very poor effect size for the main group results from the lack of difference between PPVTpre and PPVTpost. However, when calculating differential effect size, the result is much larger at d = 2.15. This suggests that taking the difference between the main group and control group into account, there is actually a large effect size, whereby the main group does not change, but the control group improves significantly.

While the effect size results are clearly very high, it is difficult to determine whether the measured increase in fatigue will have any real effects in the flight training environment. It is quite possible that the main group participants suffered some degree of cognitive impairment based on the increase in fatigue level that was measured. In any safety-conscious environment such as aviation, any degree of such impairment should at least be considered and investigated as a potential threat to effective and safe operations.

It is important to consider that the student would most likely only reach the level of fatigue that was observed near the end of their flight, and that for the majority of the flight they would be less fatigued. Therefore, this level of fatigue would only be reached after the majority of the training had been completed. From a safety perspective, having a flight instructor being present on most flights would also help to mitigate any negative effects of fatigue. In saying this, flight instructors are also subjected to fatigue potentially as much as, if not more than, the students themselves, and accidents and incidents do often occur near the end of a flight, such as during the approach and landing phases. Some of the correlations found between different variables suggest that students in earlier stages of training may be more likely to experience fatigue. However, this is countered by the fact that newer students are usually more highly supervised in the beginning stages of their training.

While participants were clearly fatigued after the training flight, they did not register any notable levels of subjective measures of impairment such as memory or concentration problems. This has potentially two implications, that while fatigue may be present, it does not create any measureable impairment amongst the participants, and also that there is a possibility that the participants were indeed affected but did not recognise it. PPVT data did not show any conclusive results either, in that PPVT did not deteriorate for the main group, but performances were worse than control group members.

There is the potential for fatigue to be a latent threat in flight training operations, in that it is not a considerable concern by itself, but has the potential to form a link in an accident causation chain, or alternately serve as a 'hole' in Reason's (2000) Swiss Cheese model. The effects of fatigue do not need to be strong enough to cause an accident directly, but they could prevent a pilot from making correct and timely decisions and actions. While normal everyday flight operations may not appear to be affected by fatigue, in the case of an emergency situation it may prove to be more of a threat. Aside from safety, the effectiveness of the training flight is another area where fatigue could prove to be a negative influence. There is the potential that the effects of fatigue could impair a student's performance so that, in addition to creating a potential safety hazard, another more visible effect is the ineffectiveness of the training flight. In a situation where a student needs to make as much progress as possible after each flight, the presence of fatigue could limit the amount of progress made, especially for longer and more intense flights.

In conclusion, while there is not likely to be a considerably high level of impairment associated with the observed increase of fatigue, it is not something that should be ignored. The effects of fatigue could also prevent a training flight from being as effective as it otherwise could be.

5.2.5. Are the results statistically significant?

Due to the relatively small sample sizes, not all of the results are statistically significant, only QFS for the main group and PPVT for the control group from the main variables. However, when considering that PPVT for the main group showed little change, and that this can be explained by points made earlier in this chapter, poor statistical significance for this measure is not as critical. Furthermore, with only 5 participants in the control group for QFS, changes need to be extremely large to be significant. If the sample size was larger, it is likely that results would be more significant. All of the important correlations considered between the different variables also have strong levels of significance, at either $p < 0.05$ or < 0.01. Overall, the statistical significance of the results seems to be satisfactory.

5.2.6. Can the results be used to set a baseline for cognitive fatigue?

The overall aim of this study is to set a baseline for cognitive fatigue in student pilots. In considering all of the data gathered, QFS seems to be the most reliable and accurate method of measuring fatigue. Therefore the change in QFS of 20% seems to be the best example to use to set such a baseline, especially given its strong correlations with other questionnaire items.

5.3. Limitations of the study

It has been acknowledged in the methodology section that some compromises had to be made in designing this study, with geographical, financial, and time restrictions effectively limiting the study to only one aviation training organisation. Time restrictions also necessitated using a shorter version of the PPVT, which while still viable, is not as ideal as the full ten-minute version. Furthermore, a small non-probability sample resulted from the number of voluntary participants available at only one ATO, and was further compromised by the limited time available for the testing, and the low rate at which testing could be completed. It is clear that the study suffered somewhat for lack of a larger sample, however with the tools and procedure now validated, there is potential for further studies to build from this work and overcome the limiting sample size. However, even in light of these compromises, the study effectively balanced all the requirements so that its integrity was not affected.

5.4. Potential for generalisation

Despite the study's limitations, there is still a strong potential for generalisation to the wider population of student pilots. It is not expected that the characteristics of different training organisations and student pilots would affect the results to a significant extent. One area where generalisation might not be as applicable is to non-certified flying schools, especially those smaller establishments with more informal operations.

5.5. Other considerations

Fatigue has been an identified causal factor in many aircraft accidents, some of which were briefly analysed in the literature review section, and it has also likely played an unseen role in many more accidents. There have been many studies of the effects of fatigue on air transport pilots, much concern about regulations surrounding fatigue, and continued attempts to prevent fatigue from causing aircraft accidents. While an airline pilot may have a heightened awareness of fatigue concerns through their employer, regulatory bodies, and the media, it is likely that the average student pilot is not so concerned with the phenomenon of fatigue. This may very well be reasonable, given the previously mentioned reasons of students not being exposed to the same risks and causes as commercial pilots. However, underestimating fatigue at the crucial early stages of flight training could have implications for the future. The data have shown that it is almost certain that a student pilot will suffer from fatigue over the course of a training flight, and the literature clearly highlights that the many effects of fatigue are varied, often difficult to detect, underestimated, and potentially deadly. Without an appreciation of the nature of fatigue, and from the lack of clear and strong effects experienced in training, there is the possibility that student pilots will continue to underestimate it throughout their training, and so develop a chronic under-appreciation of fatigue that may extend to their future aviation employment.

Chapter six – Conclusion

6.1. Overview

Flight training can be a tough environment, placing many demands on the information processing capabilities of student pilots unfamiliar with the high workload and challenging environment. However, with a lack of research aimed at student pilots, many aspects of how they cope in this environment are unknown. Fatigue has been a topic of concern in aviation for many years, and basic research has been undergoing since the early 20th century; despite this, it remains a complex topic with many different views put forward. However, one widely accepted cause of fatigue is high workload and stress, the likes of which student pilots can experience during a flying lesson. One point that has much consensus is that fatigue can have many effects, and these effects can vary greatly in occurrence and severity.

Two preliminary research tasks were undertaken to facilitate the completion of the study. The first task was to investigate the population, being student pilots training towards a fixed-wing pilot licence in New Zealand. Initial contact with the CAA lead to individual enquiries to several Part-141 authorised aviation training organisations. This identified a rough distribution of the average 799 student pilots training at the 16 different locations at the time, albeit with the understanding that numbers are prone to regular fluctuations. From the information that was gathered, a decision was made to proceed focusing on only students training at ATOs.

The next task was to establish a means of measuring fatigue as accurately and reliably as possible. Initial searching did not find any tools that had been previously validated and used in a flight training environment, and it was realised that custom tools would have to be designed and created. The first tool that was created was a two-part questionnaire which based its primary fatigue measurement from items in existing fatigue-measuring questionnaires. Many of these items were from a medical environment, and were narrowed down to ten core items, which were answered using a seven point Likert scale. When split between pre-flight and post-flight sections of the questionnaire they served as a fatigue scale [QFS] which could measure the participants' subjective evaluation of their own fatigue. In addition to the QFS, the questionnaire featured items which measured demographic criteria, and several others which measured various factors relating to fatigue.

Second, a reaction time test called PPVT was obtained from the PEBL psychology software package. PPVT [PEBL Perceptual Vigilance Task] is a version of the widely-used Psychomotor Vigilance Test [PVT], versions of which are used in many different aerospace environments. With a shorter five-minute version having been previously verified as accurate and reliable, it was decided that this test, operated from a laptop computer, would serve as an objective means of measuring reaction time, one of the negative effects of fatigue on human performance. After a brief pilot test successfully validated both tools, the main study could be commenced.

Testing began at a single large ATO, with students from non-probability convenience sampling completing the testing over a period of several weeks. The student pilots were tested both before and after a standard training flight, with each student completing both the pre-flight [QFSpre] and post-flight [QFSpost] sections of the questionnaire, as well as pre-flight [PPVTpre] and post-flight [PPVTpost] PPVT testing. While not initially a goal of the study, external factors during testing created the potential to create a small control group of student pilots, who completed as much testing of the testing as possible, but without actually completing a training flight. Twenty-one student pilots participated in the study, split between 16 in the main experimental group, and 5 in the control group. The data were initially analysed using SSPS, with favourable results suggesting continuing evaluation would make use of parametric testing. The results were addressed in four broad areas; changes between pre-flight and post-flight data for the two groups, along with correlations [Pearson's *r*] between the different variables, and evaluation of both practical [Cohen's *d*] and statistical significance [Student's t-test].

Considerable changes were identified in the QFS data, with every main group participant experiencing some degree of increased fatigue after their training flight, an average change of 1.18 points. Reasonably high levels of fatigue were measured in the main group participants, with an average deterioration of 20% as measured by QFS being the most promising result for consideration as a baseline. This contrasted even more when comparing the control group, which actually showed a decreasing fatigue level of 11%. The testing seemed to operate cohesively, while QFS and PPVT in particular did not correlate quite as well as suspected, there are reasonable explanations available for the different measurements observed.

While there may not be any major implications from the amount of fatigue that was measured, it is still worthy of consideration, particularly when considering the potential of fatigue as a safety hazard, and the possible follow-on effects into other areas of aviation.

6.2. Key outcomes

This study has added to the overall body of knowledge about fatigue in aviation, and particularly in flight training, through four main results.

First, many previous studies note a lacking in means of measuring fatigue, so the successful design and validation of the questionnaire, particularly QFS, is a great result. It can be used to measure fatigue in student pilots, and also potentially adapted for use in other areas of aviation. Versions of the PPVT have been in use for many years in many different settings, and so its adaption for this study entailed little in the way of modifications or redesign.

Second, the development and execution of a testing procedure for measuring fatigue levels in student pilots is quite possibly one of the first times this has been accomplished. While arguably not exactly a challenging or impressive task in itself, it is one that has had comparatively little attention, with studies of fatigue instead mostly focusing on qualified military and civilian pilots. As a result, this study can possibly act as a catalyst for further study in the area, and leaves many new aspects that can be investigated, as will be further outlined in section 6.3.

Third, this study has produced some of the only data available on cognitive fatigue levels in student pilots, as well as results that are highly significant in both a statistical and practical sense. The different tools and measures relate well, and the outcomes are both interesting, and have strong links to the literature, reinforcing previous knowledge on fatigue in aviation environments. There are many correlations between workload, fatigue, level of preparedness to undergo a training flight, and performance impairments. Most of these relationships are to be expected, and predicted by the literature review from fatigue studies in other environments. However, as in the previous point outlined, the fact that obtaining this specific information about student pilots has not been previously accomplished lends increased merit to the achievement.

Finally, the goal of setting a baseline of cognitive fatigue in student pilots has been achieved, with a figure of 20% as indicated by QFS being a highly significant result which also correlates well to other factors.

6.3. Areas for future research

With little other research in the area, there is great potential for future investigation. One possibility that was considered when designing this study was to incorporate a judgement of the student's performance by their flight instructor. A flight instructor is in an ideal position to make this judgement, as they are less likely to underestimate any performance decreases, and would have probably flown with a particular student several times, having a good idea of how they perform. Furthermore, having likely completed many training flights with different students, flight instructors will probably be in a better position to detect changes in performance compared to a student who is still accommodating to a new environment and new lesson material.

While this study was reasonably limited in its sampling criteria, there is the possibility to expand the sample size by scaling back the researcher's involvement in the testing. For example, now that the questionnaire has been successfully validated and provided reliable results, it is possible to only use the questionnaire and forego the PPVT; the testing could then be self-administered by participants. In this case, it would only be necessary to send questionnaires directly to training organisations and have the students complete and return them. This could greatly increase the potential sample size, and as the questionnaire arguably delivered the best results it would not compromise the integrity of the research.

Furthermore, with a thorough solicitation of non-certified training organisations, possibly through organisations such as Flying NZ, the study could then be further expanded to include other areas of flight training. This would likely be incorporated with self-administered testing to avoid the difficulties of travelling to many different locations.

The same techniques could also be applied in an investigation of rotary-wing flight training.

Another interesting use of the questionnaire would be to have a student and their instructor both complete separate questionnaires for the same training flight, and also possibly comment on each other's apparent fatigue levels.

A longitudinal study is also another possibility, whereby the same students are tested multiple times at different stages of their training. This could require an expansion of the questionnaire, possibly by rewording the core items, or adding more, to ensure the same items are not repeated too soon so as to create familiarity.

A final possibility is to administer the same questionnaire to pilots conducting different kinds of flight operations, and then comparing results to see which areas might be most at risk. This could require a re-wording of items that directly relate to the 'training' nature of a flight so as to create a generic questionnaire that would be equally applicable to different flight operations.

References

Bennett, S. A. (2003). Flight crew stress and fatigue in low-cost commercial air operations - an appraisal. *International Journal of Risk Assessment and Management, 4*(2), 207-231.

Beurskens, A. J. H. M., Bültmann, U., Kant, I., Vercoulen, J. H. M. M., Bleijenberg, G., & Swaen, G. M. H. (2000). Fatigue among working people: Validity of a questionnaire measure. *Occupational and Environmental Medicine, 57*, 353–357.

Bor, R., Field, G.,& Scragg, P. (2002). The mental health of pilots: An overview. *Counselling Psychology Quarterly, 15*(3), 239-256.

Bourgeois-Bougrine, S., Carbon, P., Gounelle, C., Mollard, R., & Coblentz, A. (2003). Perceived fatigue for short- and long-haul flights: A survey of 739 airline pilots. *Aviation, Space, and Environmental Medicine, 74*(10), 1072-1077.

Caldwell, J. A., & Caldwell, J. L. (2003). *Fatigue in aviation: A guide to staying awake at the stick*. Aldershot, England: Ashgate.

Caldwell, J., Caldwell, J. L., Brown, D., Smythe, N., Smith, J., Mylar, J., ... & Schroeder, C. (2003). *The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots*. Brooks City-Base, TX: United States Air Force Research Laboratory.

Caldwell, J. A. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease, 3*(2), 85-96.

Caldwell, J. A., Mallis, M. M., Caldwell, J. L., Miller, J., Paul, M., & Neri, D. (2009). Fatigue countermeasures in aviation. *Aviation, Space, and Environmental Medicine, 80*(1), 29-59.

Campbell, R. D., & Bagshaw, M. (1991). *Human performance and limitations in aviation*. Oxford: Blackwell.

Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D., & Wallace, E. P. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research, 37,* 147–153.

Civil Aviation Authority. (2000). Fatigue. *CAA News, Nov/Dec*. Wellington, New Zealand: Author.

Civil Aviation Authority. (n.d.). *Part 141 training organisation holders*. Retrieved 14 November, 2012, from: http://www.caa.govt.nz/Script/AvDocClientList.asp?Details_ID=689

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Mahwah, NJ: Lawrence Erlbaum.

Crawford, J., & Neal, A. (2006). A review of the perceptual and cognitive issues associated with the use of head-up displays in commercial aviation. *The International Journal of Aviation Psychology, 16*(1), 1-19.

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

Dinges, D. (2012). *Psychomotor vigilance self test on the International Space Station*. Retrieved 16 October 2012 from: http://www.nasa.gov/mission_pages/station/research/experiments/Reaction_Self_Test.html

Dittner, A. J., Wessely, S. C., & Brown, R. G. (2004). The assessment of fatigue: A practical guide for clinicians and researchers. *Journal of Psychosomatic Research, 56*, 157-170.

Ewing, R. (2003). *Aviation medicine and other human factors for pilots*. Christchurch, New Zealand: Old Sausage.

Fletcher, A., Lamond N., van den Heuvel C. J., & Dawson D. (2003). Prediction of performance during sleep deprivation and alcohol intoxication using a quantitative model of work-related fatigue. *Sleep Research Online, 5*(2), 67-75.

Flying New Zealand. (n.d.) *About Flying New Zealand*. Retrieved 14 November, 2013, from: http://www.flyingnz.co.nz/about.html

Fuchs, T., & Burgdorf, J. (2008). Replication and pedagogy in the history of psychology IV: Patrick and Gilbert (1896) on sleep deprivation. *Science and Education, 17*(5), 511-524.

Gartner, W. B., & Murphy, M. R. (1976). *Pilot workload and fatigue: A critical survey of concepts and assessment techniques*. Washington D.C.: National Aeronautics and Space Administration.

Godwin, P.D. (2006*). Human factors & pilot performance*. Cranfield, England: Air Pilot Publishing.

Goode, J. H. (2003). Are pilots at risk of accidents due to fatigue? *Journal of Safety Research, 34*(3), 309-313.

Gundel, A., Drescher, J., Maaβ, H., Samel, A., & Vejvoda, M. (1995). Sleepiness of civil airline pilots during two consecutive night flights of extended duration. *Biological Psychology, 40*(1), 131-141.

Hunter, D. R. (2005). Measurement of hazardous attitudes among pilots. *The International Journal of Aviation Psychology, 15*(1), 23-43.

Jackson, C. A., & Earl, L. (2006). Prevalence of fatigue among commercial pilots. *Occupational Medicine, 56*(4), 263-268.

Lamond N., & Dawson, D. (1999). Quantifying the performance impairment associated with fatigue. *Journal of Sleep Research, 8*, 255-62.

Lee, K. E., Hicks, G., & Nino-Murcia, G. (1993). Validity and reliability of a scale to assess fatigue. *Psychiatry Research, 36*, 291–298.

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behaviour Research Methods, Instruments, & Computers, 36(*2), 339-346.

Michielsen, H. J., de Vries, J., & van Heck, G. L. (2003). Psychometric qualities of a brief self-rated fatigue measure: The fatigue assessment scale. *Journal of Psychosomatic Research, 54*, 345-352.

Michielsen, H. J., de Vries, J., van Heck, G. L., van de Vijver, F. J. R., & Sijtsma, K. (2004). Examination of the dimensionality of fatigue: The construction of the fatigue assessment scale (FAS). *European Journal of Psychological Assessment, 20*(1), 39-48.

Miyagi, M. (2005). *Serious accidents and human factors: Breaking the chain of events leading to an accident: Lessons learned from the aviation industry.* Chichester, UK: John Wiley & Sons.

Montgomery, V. L. (2007). Effect of fatigue, workload, and environment on patient safety in the pediatric intensive care unit. *Pediatric Critical Care Medicine, 8*(2), 11-16.

Mueller, S. T. (2010). *The Psychology Experiment Building Language, version 0.12*. Retrieved September 12, 2012, from: http://pebl.sourceforge.net.

National Space Biomedical Research Institute. (2010). *520-day Mars mission simulation: U.S study looks at impact of stress and fatigue on performance*. Retrieved March 14, 2013, from: http://www.nsbri.org/newsflash/indivArticle.asp?id=454&articleID=90

National Transportation Safety Board. (1973). *Aircraft accident report: Eastern Air Lines, Inc., L-1011, N310EA, Miami, Florida, December 29, 1972*. Washington D.C.: Author.

National Transportation Safety Board. (1980). *Aircraft accident report: Air New England, Inc., deHavilland DHC-6-300, N383EX, Hyannis, Massachusetts, June 17, 1979*. Washington D.C.: Author.

National Transportation Safety Board. (2000). *Aircraft accident report: Controlled flight into terrain, Korean Air Flight 801, Boeing 747-300, HL7468, Nimitz Hill, Guam, August 6, 1997*. Washington D.C.: Author.

National Transportation Safety Board. (2010). *Loss of control on approach, Colgan Air, Inc., operating as Continental Connection Flight 3407, Bombardier DHC-8-400, N200WQ, Clarence Center, New York, February 12, 2009*. Washington D.C.: Author.

Petrilli, R. M., Roach, G. D., Dawson, D., & Lamond, N. (2006). The sleep, subjective fatigue, and sustained attention of commercial airline pilots during an international pattern. *Chronobiology International, 23*(6), 1357-1362.

Powell, D. M. C., Spencer, M. B., Holland, D., Broadbent E., & Petrie, K. J. (2007). Pilot fatigue in short-haul operations: Effects of number of sectors, duty length, and time of day. *Aviation, Space, and Environmental Medicine, 78*(7), 698-701.

Reason, J. (2000). Human error: Models and management. *BMJ, 320*, 768-770.

Robson, D. (2008). *Human being pilot: Human factors for aviation professionals*. Cheltenham, Australia: Aviation Theory Centre.

Rosekind, M. R., Graeber, R. C., Dinges, D. F., Connell, L. J., Rountree, M. S., Spinweber, C. L., & Gillen, K. A. (1994). *Crew factors in flight operations IX: Effects of planned cockpit rest on crew performance and alertness in long-haul operations (NASA Technical Memorandum 108839)*. Moffett Field, CA: NASA Ames Research Center.

Ritter, R. D. (1993). 'And we were tired': Fatigue and aircrew errors. *Aerospace and electronic systems magazine, IEEE, 8*(3), 21-26.

Samkoff, J. S., & Jacques, C. H. M. (1991). A review of studies concerning effects of sleep deprivation and fatigue on residents. *Academic Medicine, 66*(11), 687-693.

Schwartz, J. E., Jandorf, L., & Krupp, L. B. (1993). The measurement of fatigue: A new instrument. *Journal of Psychosomatic Research, 37*, 753–762.

Sharpe, M., & Wilks, D. (2002). *Fatigue, 325*(7362), 480-483.

Signal, L., Ratieta, D., & Gander, P. (2006). *Fatigue management in the New Zealand aviation industry*. Canberra: Australian Transportation Safety Bureau.

Smith, A. P., Borysiewicz, L., Pollock, J., Thomas, M., Perry K., & Llewelyn, M. (1999). Acute fatigue in chronic fatigue syndrome patients. *Psychological Medicine, 29*(2), 283-290.

Stokes, A., & Kite, K. (1994). *Flight stress: Stress, fatigue, and performance in aviation*. Aldershot, England: Ashgate.

Tabacknick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics (4th ed)*. Boston, MA: Allyn & Bacon.

Tan, E. M., Sugiura, K., & Gupta, S. (2002). The case definition of chronic fatigue syndrome. *Journal of Clinical Immunology, 22*(1), 8-12.

Trollip, S. R., & Jensen, R. S. (1991). *Human factors for general aviation*. Englewood, CO: Jeppesen Sanderson.

Wilkinson, R. T., & Houghton, D. (1982). Field test of arousal: A portable reaction timer with data storage. *Human Factors, 24*, 487-493.

Wilson, G. F., Russell, C., & Caldwell, J. (2006). *Performance and psychophysiological measures of fatigue effects on aviation related tasks of varying difficulty*. Patterson Air Force Base, OH: United States Air Force Research Laboratory.

Yen, J., Hsu, C., Yang, H., & Ho, H. (2009). An investigation of fatigue issues on different flight operations. *Journal of Air Transport Management, 15*(5), 236-240.

Appendix A: Sample questionnaire

Ab-initio flight training questionnaire

IMPORTANT INFORMATION FOR PARTICIPANTS - PLEASE READ

Researcher
Nicholas J Ashley, BAvMan, Massey University School of Aviation
nickashley172@gmail.com, 021 0261 2814

Supervisor
Jose D Perezgonzalez, PhD, Massey University School of Aviation
J.D.Perezgonzalez@massey.ac.nz, 06 356 9099 x5326

Overview of Study
To measure the potential change in selected variables among participants after completing an ab-initio training flight

Ethics Disclaimer
This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher named above is responsible for the ethical conduct of this research
If you have any concers about the conduct of this research that you wish to raise with someone other than the research , please contact Professor John O'Neill, Director (Research Ethics), telephone 06 350 5249, e-mail humanethics@massey.ac.nz

Participant rights and privacy

Please remember that participating in this study is completely voluntary and you may choose to stop at any time

All information collected as part of this study will remain confidential and will not be traced back to you

Please DO NOT record any unsolicited personal information such as your name on the questionnaire, this is to protect your privacy

Pre-flight questionnaire - Instructions

This questionnaire is to be completed immediately BEFORE the training flight, but AFTER any associated pre-flight checks, planning, or briefing commences. That is, it should be completed as soon as possible before entering the aircraft to begin the practical stage of the lesson.

There are two sections, the first is an overview of personal information which will be used to assist in analysing the results. Please complete this by placing a tick in the box or by writing in the space provided, as appropriate. The second section consists of several questions which are to be answered by circling the most appropriate number. Please read the initial sentence in bold before you answer the questions.

For example, if you had to respond to the statement: "I find flying is a challenge", and you agreed with this, but not completely, circle 5 or 6 such as in the example below:

	Strongly disagree						Strongly agree
I find flying is a challenge	1	2	3	4	(5)	6	7

pg 1

Pre-flight questionnaire - Section 1

Age:________

Gender:
Male □
Female □

Total flight time:__________

Hours of sleep in the past 24 hours:__________

Usual number of hours of sleep each night:__________

Training aircraft type to be used for the upcoming flight:________________________

Type of cockpit display used:
Analogue □
Technically enhanced, or 'glass cockpit' □

Name of lesson to be completed e.g. "Medium turns":__________________________

PPL theory passes:
Air law □
English proficiency □
Flight radio □
Human factors □
Meteorology □
Navigation □
Technical knowledge □

Please provide details of ANY medication consumed within the past 24 hours:

Name/type:____________________ Time consumed:__________

Name/type:____________________ Time consumed:__________

Please provide details of the MOST RECENT caffeinated beverage (if any) consumed today:

Name/type:____________________ Time consumed:__________

Please indicate the time of the MOST RECENT meal consumed today (if any, and EXCLUDING light snacks):

Time consumed:__________

pg 2

Pre-flight questionnaire - Section 2

Please think about your upcoming training flight and indicate the extent to which you agree or disagree with each statement below:

	Strongly disagree						Strongly agree
Mentally, I am well prepared to undergo the flight	1	2	3	4	5	6	7
There are new concepts or techniques featuring in the flight	1	2	3	4	5	6	7
I feel energetic	1	2	3	4	5	6	7
I feel more forgetful than normal	1	2	3	4	5	6	7
I can concentrate very well	1	2	3	4	5	6	7
I have an overall feeling of tiredness	1	2	3	4	5	6	7
I feel mentally exhausted	1	2	3	4	5	6	7
The flight will mostly be revision of concepts and techniques that I am already familiar with	1	2	3	4	5	6	7

Thank you for completing this questionnaire. Your time is appreciated.

Please go back and ensure you have fully completed every item.

You may then hand in your completed questionnaire.

pg 3

Post-flight questionnaire - Instructions

This questionnaire is to be completed immediately AFTER the training flight, but BEFORE any associated debriefing commences. It consists of additional questions which are to be answered by circling the most appropriate number.
Please read the initial sentence in bold before you answer the questions.
For example, if you had to respond to the statement: "I find flying is a challenge", and you agreed with this, but not completely, circle 5 or 6 like in the example below:

	Strongly disagree						Strongly agree
I find flying is a challenge	1	2	3	4	(5)	6	7

pg 1

Post-flight questionnaire

Please indicate the extent to which you agree or disagree with each statement below:

	Strongly disagree					Strongly agree	
I am pleased with the outcome of the flight	1	2	3	4	5	6	7
I was well prepared to undergo the flight	1	2	3	4	5	6	7
Mentally, I feel in good condition	1	2	3	4	5	6	7
I find it difficult to concentrate	1	2	3	4	5	6	7
I can think clearly	1	2	3	4	5	6	7
I feel relaxed	1	2	3	4	5	6	7
I feel worn out	1	2	3	4	5	6	7

Please consider the events over the course of the flight you have just completed and indicate the extent to which you agree or disagree with each statement:

	Strongly disagree					Strongly agree	
There was a high workload throughout the flight	1	2	3	4	5	6	7
I found it easy to recall important points and concepts relating to the training flight	1	2	3	4	5	6	7
My performance gradually decreased throughout the flight	1	2	3	4	5	6	7
My performance sharply decreased near the end of the flight	1	2	3	4	5	6	7
I would be able to perform at the same level of proficiency on another similar training flight to be commenced now	1	2	3	4	5	6	7
I would be able to perform at a higher level on another similar training flight to be commenced now	1	2	3	4	5	6	7
I became easily distracted at times near the end of the flight	1	2	3	4	5	6	7
I made small mistakes or forgot things at the start of the flight	1	2	3	4	5	6	7

	Strongly disagree						Strongly agree
I made small mistakes or forgot things near the end of the flight	1	2	3	4	5	6	7
My flying ability improved over the course of the flight	1	2	3	4	5	6	7
I was able to maintain a focus on the requirements of the training flight throughout its duration	1	2	3	4	5	6	7

Please consider your current ability to cope with the demands of a training flight and indicate the extent to which you agree or disagree with each statement below:

	Strongly disagree						Strongly agree
I am well prepared to complete another training flight commencing now	1	2	3	4	5	6	7
I would need to have some time to rest and recover before completing another similar flight today	1	2	3	4	5	6	7

Thank you for completing this questionnaire. Your time is appreciated.

Please go back and ensure you have fully completed every item.

You may then hand in your completed questionnaire.

Appendix B: Questionnaire items

Code	Item text
PRE01	Mentally, I am well prepared to undergo the flight
PRE02	There are new concepts or techniques featuring in the flight
PRE03	I feel energetic
PRE04	I feel more forgetful than normal
PRE05	I can concentrate very well
PRE06	I have an overall feeling of tiredness
PRE07	I feel mentally exhausted
PRE08	The flight will mostly be revision of concepts and techniques that I am already familiar with
POST01	I am pleased with the outcome of the flight
POST02	I was well prepared to undergo the flight
POST03	Mentally, I feel in good condition
POST04	I find it difficult to concentrate
POST05	I can think clearly
POST06	I feel relaxed
POST07	I feel worn out
POST08	There was a high workload throughout the flight
POST09	I found it easy to recall important points and concepts relating to the training flight
POST10	My performance gradually decreased throughout the flight
POST11	My performance sharply decreased near the end of the flight
POST12	I would be able to perform at the same level of proficiency on another similar training flight to be commenced now
POST13	I would be able to perform at a higher level on another similar training flight to be commenced now
POST14	I became easily distracted at times near the end of the flight
POST15	I made small mistakes or forgot things at the start of the flight
POST16	I made small mistakes or forgot things near the end of the flight
POST17	My flying ability improved over the course of the flight
POST18	I was able to maintain a focus on the requirements of the training flight throughout its duration
POST19	I am well prepared to complete another training flight commencing now
POST20	I would need to have some time to rest and recover before completing another similar flight today

Underlined items, PRE03-PRE07 and POST03-POST07, are components of QFS