# Physiological Arousal, Information Processing, Performance and Expertise in Expected and Unexpected Abnormal Flight Events: An Empirical Investigation.

Lana Kinney

(6088157)

*Thesis submitted in partial fulfilment of the requirements for the degree of*

*Master of Science*

University of Otago, Dunedin, New Zealand

March 2018

## Abstract

Aviation accident reports indicate that preventable incidents are developing into tragedies with pilots responding incorrectly to well-trained events e.g. engine failures. Recent research suggests that startle (an autonomic response to an acute stimulus with a sudden onset), following unexpected abnormal flight events is impacting pilot performance, leading to accidents. The present study was designed to investigate whether a simulated unexpected abnormal flight event can lead to startle. Information processing and performance differences between expected and unexpected flight events were also measured. Furthermore, the influence of expertise on arousal, information processing and performance in these events was investigated. Two studies were conducted. The first study employed university students recruited through the University of Otago Psychology Database. The second study employed general aviation pilots recruited through social media advertising. Students and pilots flew a series of flights in a fixed-base flight simulator including four experimental flights which included an unexpected or an expected, engine failure or aerodynamic stall. During the flights, heart rate, eye-tracking, and flight data were recorded. Increased heart rate and larger pupil dilation during the unexpected engine failure indicated the presence of startle in pilots. During the unexpected engine failure pilots showed a disrupted information processing strategy that indicated attentional tunnelling. Whereas, during the unexpected stall the information processing patterns indicated lack of recognition. During the unexpected events performance was impaired when compared to the expected events. However, poor performance was not associated with higher levels of arousal. In a third comparative study, data from novice (university students), intermediate (student and private licenced) and expert (commercial licenced) pilots were compared to investigate the effects of expertise. Information processing, arousal, and performance did not differ significantly over the three levels of expertise. This research supports a recently formulated theory on startle and surprise and has implications for successful training.

**Acknowledgements**

Foremost, I would like to thank the students and the pilots who sacrificed their time to participate in this research. I would also like to thank in advance the examiners for all the time they will be needing give up to grade this colossal Master's thesis.

I would like to specially acknowledge my supervisor Professor David O'Hare. Professor. O'Hare gave me continuous support and a much needed keen eye throughout the entire thesis process. He allowed me to work independently which helped me to learn and develop my research skills. However, he was always there to steer me in the right direction, proof read, and show me relevant literature he may have stumbled across.

Furthermore, thank you to my friends for listening, taking me outside for walks, and sending me care packages. I would like to thank my peers in the Cognitive Engineering and Decision Making Laboratory, I owe you all my sanity many times over, and thank you for all the laughs. Finally to my family, Kaye, Martin, Luke, and my partner Jesse, thank you for your unfailing support, your words of wisdom, and your constant encouragement throughout my years of study and my postgraduate degree.

## Table of Contents

## List of Tables



## List of Figures

*"Aviation in itself is not inherently dangerous. But to an even greater degree than the sea, it is terribly unforgiving of any carelessness, incapacity or neglect".*

*— Captain A. G. Lamplugh circa 1930s.*

**Stress and Emergency Situations**

The influence of stress during emergency situations on human behaviour and performance has been of interest throughout the history of psychology, particularly for governmental agencies and the military (Bourne Jr & Yaroush, 2003; Driskell & Salas, 2013; Lazarus, Deese, & Osler, 1952; Staal, 2004). Today's modern world contains vastly complex and highly technological systems throughout industries such as aviation, maritime, petroleum, mining, rail, and nuclear. These systems are now more complex than ever before, for example, the Airbus A380 planes can now carry more than 800 passengers, and modern oil tankers can now get up to 1000 feet long. Emergency events can occur suddenly and unexpectedly, and fast, accurate decisions are needed from human operators who are commonly under extreme stress (Driskell & Salas, 2013; Proctor & Van Zandt, 2018; Woodson, Tillman, & Tillman, 1992). Due to the complexity of the systems, errors made, can lead to instant and destructive consequences, therefore research into the impact of stress on performance is of great importance.

Driskell and Salas (2013) defined stress as "A process by which certain environmental demands evoke an appraisal process in which perceived demand exceeds resources, and results in undesirable physiological, psychological, behavioural, or social outcomes" (p.6). Acute stress was defined as an affect which is "sudden, novel, intense and of relatively short duration, disrupts goal-oriented behaviour, and requires a proximate response" (Driskell & Salas, 2013, p.6). Startle or surprise can lead to acute stress in complex systems e.g. abnormal flight events (Landman, Groen, van Paassen, Bronkhorst, & Mulder, 2017a). The presence and effects of startle due to unexpected abnormal flight events in general aviation is the focus of this present study.

**Stress and Performance**

The function of stress is to recruit and allocate mental resources to enable an individual to respond optimally in demanding situations (Landman et al., 2017a). However, previous research on the effects of psychological stress on behaviour shows that aspects of stress (e.g. impaired top down performance, changes in attentional control, physiological arousal, and performance rigidity) can be detrimental to performance (Dismukes, Goldsmith, & Kochan, 2015; Eysenck, Payne, & Derakshan, 2005; Landman et al., 2017a; Wickens, Stokes, Barnett, & Hyman, 1993). Early research has focused on simple motor and perception tasks. These investigations have consistently shown that stressful conditions degrade performance (Bourne Jr & Yaroush, 2003; Jensen, 1995; Staal, 2004). More recent findings show that higher order cognitive processes tend to be more sensitive to psychological stress in comparison to simple perceptual motor skills (Bourne Jr & Yaroush, 2003; Staal, 2004). Kivimäki and Lusa (1994) found a negative relationship between stress and controlled task-focused thinking, when investigating fire fighter's performance and cognitive functioning under severe stress. Consistent with this finding, stress has been shown to cause a failure to think beyond prescribed procedures, and reduce problem solving capacities during complex or ambiguous situations (Broadbent, 1971). Training that improved cognitive processing during stressful situations has been found to be the most helpful, leading to the least degraded performance in stressful operations (Kivimäki & Lusa, 1994). In summary, whilst the cognitive impact of stress impairs performance, training targeting processing during stressful situations can mediate this problem to some extent.

Stressful and threatening stimuli have been found to monopolize attentional resources when they are perceived (Staal, 2004). Therefore, stress can lead to a reduction in peripheral cue utilisation and narrowing of the attentional field (Baddeley, 1972; Combs & Taylor, 1952; Easterbrook, 1959; Staal, 2004). Stimuli that are perceived as salient or important by an individual, generally receive preferential attentional processing (Staal, 2004). Stressful and threatening stimuli tend to be perceived as salient and important and therefore monopolise more attentional resources (Staal, 2004). Concordant with this, Combs and Taylor (1952) found that when translating sentences into code, participants took longer when the sentences contained threats. Furthermore, military and open sea diving research shows that with increasing amount of danger, stress leads to attentional narrowing and a decrement in performance (Baddeley, 1972). This can sometimes be exacerbated to the point where individuals will completely abandon operation of the control systems (Baddeley, 1972). Expertise has been shown to mediate this effect (Baddeley, 1972). Stress can also cause attentional processing to become disordered due to unusually high levels of arousal and vigilance (Staal, 2004). This condition often results in an indiscriminate search, fast disordered attentional shifting, and a reduction in the number and quality of other courses of action considered (Staal, 2004). Streufert and Streufert (1981) reported that with increased stress due to time pressure participants made faster decisions. As well as this, there was a decrease in information utilisation and search behaviour with the increased stress caused by time pressure (Streufert & Streufert, 1981). This research concluded that stress affects performance by impacting an individual's breadth of attention and ability to logically and calmly process information.

Working memory refers to the mental ability of temporarily retaining, updating and manipulating information (Baddeley, 2003; Jiang & Rau, 2017). Working memory is thought to be the basis of crucial cognitive processes including planning, logical thinking, and problem solving (Porcelli et al., 2008). Research shows that different stressors, e.g. noise, social stress, or time pressure impair the ability to perform working memory tasks, and reduces the working memory capacity (Gomes, Martinho Pimenta, & Castelo Branco, 1999; Jiang & Rau, 2017; Staal, 2004; Wickens et al., 1993). For example, Recently Jiang and Rau (2017) found that the Trier social stress test (public speaking and mental arithmetic in front of an audience) increased response time on an n-back task. However, Duncko, Johnson, Merikangas, and Grillon (2009) found that in response to a cold pressor test, participants had a reduced reaction time, even in trials with greater cognitive load. However, they also found that those in the cold pressor condition tended to show higher number of false recognitions in target absent trials, but no such trend was found in target present trials. The authors suggest that target absent trials represent a slightly higher cognitive challenge than target present trials (Duncko et al., 2009). This improvement in reaction time could be helpful in stressful and threatening situations if it’s produced by faster information processing (Duncko et al., 2009). However, the authors noted that these findings may have only occurred because the cold pressor task did not increase salivary cortisol level and therefore may not be causing a high level of stress (Duncko et al., 2009). Therefore, stress has been shown to significantly affect working memory processes which underlie complex tasks. However, whether the effect is detrimental or beneficial may depend on the task.

Decision making refers to the cognitive process of choosing a particular course of action, strategy, or belief among other options on the basis of their subjective value (Vaidya & Fellows, 2017). Decision making underlies flexible and goal-orientated behaviour (Vaidya & Fellows, 2017). Research has shown that stress can impair decision making (Lehner, Seyed-Solorforough, Connor, Sak, & Mullin, 1997; Staal, 2004; Streufert & Streufert, 1981; Wickens et al., 1993). Stress tends to cause decision making to become more rigid, where fewer options are considered or processed (Broadbent, 1971; Staal, 2004; Streufert & Streufert, 1981). Lehner et al. (1997) investigated control team decision making under stress. They found that with increasing time stress, teams used more familiar but less effective decision making strategies than they were trained to use (Lehner et al., 1997). Wickens et al. (1993) investigated the idea that anxiety provoking situations may negatively influence the quality of aviation pilot decision making. Wickens et al. (1993) found that various stressors (noise, time pressure, financial risk) impaired pilot decision making and optimality in a microcomputer-based simulation of pilot decision tasks when compared to a non-stressed group. These pilot decision tasks varied in spatial processing, working memory processing, and knowledge. In summary it is well accepted that stress can impair cognitive abilities such as attention, memory and decision making, and this can affect performance on both basic and higher order tasks e.g. aviation pilot performance.

**The Startle Response and Performance**

In the early 20th century there was a need to expand the number of aviators in America due to the impending war. For recruitment purposes, the council of the American Psychological Association established a Committee on Psychological Problems of Aviation which developed a battery of 10 psychological tests predicting performance in flight training. The best predictors of good performance were emotional stability, perception of tilt, and mental alertness (Koonce, 1984). Emotional stability referred to a test of an individual's response to sudden excitation, typically from a loud noise (Koonce, 1984). This kind of testing has been phased out, however the startle reflex appears to be a major contributor to pilot malfunction in modern abnormal flight events (Landman et al., 2017a; Li, Baker, Grabowski, & Rebok, 2001; Rivera, Talone, Boesser, Jentsch, & Yeh, 2014).

The startle reflex is an autonomic response to an unexpected auditory, visual, or tactile stimulus with an abrupt onset (Davis, 1984). The physical reflexive startle response begins with muscle contractions (eye-blinks, head ducks, and crouched shoulders) followed by a quick movement away from the stimulus (Davis, 1984; Rivera et al., 2014). The startle response includes the startle reflex as well as emotional and cognitive responses where attentional resources are oriented towards the startling stimulus, (Davis, 1984). Research on fear conditioning has shown that when startle occurs in the presence of perceived threat the response can become exacerbated leading to what is known as fear potentiated startle (Bradley, Moulder, & Lang, 2005; Eysenck et al., 2005). This refers to when startle initiates a sympathetic nervous system 'fight or flight' response, including changes in blood pressure, increased heart rate (tachycardia), as well as neuroendocrine hormone release (LeDoux, 2003). Fear potentiated startle has also been found to be related to heart bradycardia, which is associated with the defensive mechanism of freezing (Bradley, Moulder, Lang, 2005).

Startle and surprise in a complex system can lead to acute stress (Landman et al., 2017a). Acute stress is when an individual appraises the situation to be threatening, taxing or exceeding ones resources (Landman et al., 2017a; Staal, 2004). Similar to other stress types, research shows that stress from startle can impair information processing (Eysenck et al., 2005; Thackray & Touchstone, 1983), as well as reduce working memory capacity (Bradley et al., 2005). Restriction of cue sampling and a narrowing of perceptive field (also known as tunnel vision) also occurs during startle which can lead to decisional errors (Driskell & Salas, 2013; May & Rice, 1971; Staal, 2004).

Early research investigated the impact of startle on simple perceptual motor tasks. May and Rice (1971) found that startle due to a loud pistol shot disrupted performance on a simple motor task (time required to press a button) and increased response time. Sternbach (1960) showed that individuals with a larger physiological startle reaction to an unexpected pistol shot had slower reaction times. Carlsen et al. (2008) investigated the effects of startle caused by a loud noise on a simple go or no go task (pressing or not pressing a button depending on colour shown). Pre-motor response time was similar in the startle and non-startle trials, however participants made more errors in the startle trials. This research provides evidence that startle can negatively impact ongoing cortical processes during simple tasks.

The research into the impact of surprise and startle on performance has shown variable findings. Research has shown that response impairment for simple reactions after startling stimuli is around 1-3 seconds (Thackray, 1965; Thackray & Touchstone, 1983). Whereas when the task is more complex and requires continuous psychomotor control, startle has been shown to cause maximum disruption during the first 3 seconds, however the disruption continued for another 10 seconds following the stimulus (May & Rice, 1971; Thackray & Touchstone, 1970). Other research suggests that after a startling stimulus information processing can be impaired for up to 60 seconds (Thackray, 1988; Thackray & Touchstone, 1970; Vlasak, 1969; Woodhead, 1969). Vlasak (1969) found that following startle, participants were impaired on a mental subtraction task for the first 30 seconds. Woodhead (1969) observed that on a continuous symbol matching task, there were decrements in performance ranging from 17-31 seconds after startle induced by a loud reproduced sonic bang (Woodhead, 1969). Additionally, when participants respond after a startling stimulus, a larger range of response times are found (Thackray, 1965; Thackray & Touchstone, 1983). Thackray (1988) also observed that response times were generally unaffected by startle however more incorrect responses were made in the startle group compared to the non-startle group, representing information processing errors. In conclusion, startle appears to negatively impact information processing for differing amounts of time. This appears to be dependent on the task, with complex tasks being more disrupted compared to simple perceptual motor tasks.

**Research into Stress, Startle and Pilot Performance**

Pilots practice abnormal flight events in training which are unlikely to occur but are important to be prepared for (Casner, Geven, & Williams, 2012). These abnormal events include aerodynamic stalls, engine failures, and hazardous weather encounters (Casner et al., 2012). Standard operating procedures are devised by airline companies to teach uniform responses to abnormal emergency aviation events and pilots practice these thoroughly (Casner et al., 2012). However, training for abnormal flight events may not always be effective. America's National Transportation Safety Board (NTSB) aviation accident reports reveal that pilots are sometimes responding inappropriately to what should be well-practiced emergency events (NTSB; 1995, 2004, 2010a, 2010b). For example; on July 13, 2003 Air Sunshine Flight 527 ditched into the Atlantic Ocean just 7.35 nautical miles from an airport in the Bahamas; two passengers died from their injuries. The cause of the crash was an in-flight failure of the right engine and the pilot's failure to manage the airplane's performance after the engine failed. The pilot had 8000 total flying hours and engine failures are extensively practiced abnormal events (NTSB, 2004). Additionally it has been shown that pilots have trouble recovering from aerodynamic stalls when they have not reviewed the recovery procedures immediately beforehand (Ledegang & Groen, 2015). Incidents such as the Air Sunshine Flight are too common and suggest some shortcomings in the pilot training paradigm for abnormal flight events. Concordant with this faulty pilot judgement is accepted as a leading cause of pilot accidents (Jensen, 1995; Landman et al., 2017a; Li et al., 2001; Rivera et al., 2014; Wickens et al., 1993). Accident reports such as this have stimulated research into stress and pilot performance.

Research at the Aviation Psychology Program of the Army Air Force in 1947 investigated the use of stress as a psychometric device for personnel selection. Specifically they looked at the effect of verbal threat and distraction on performance on various psychomotor tasks such as steadiness and aiming (Melton, 1947). Stress produced a small decrement in performance, which was worse when the stress test occurred first in the battery (Melton, 1947). The Federal Aviation Association (FAA) has stated that in high stress situations such as emergency events, pilot information scan can be severely reduced even to the point that a pilot is only focusing on one instrument (FAA, 1988). Thus, there needs to be an increase in the focus of developing the training systems to promote better acquisition of complex skills and emergency situation stress resilience i.e. generalisability to a wider range of emergency conditions (Casner et al., 2012; Salas, Bowers, & Rhodenizer, 1998).

Emergency situations are considered to be dynamic as each reaction and decision made by the responder will directly influence the resulting situation and following sequence of events (Bourne Jr & Yaroush, 2003). Emergencies are generally dependent on time, are complex, and can be stressful due to creating intense psychological demands on responders (Bourne Jr & Yaroush, 2003). There is evidence that impaired pilot performance in many incidents and accidents is due to the development of fear potentiated startle in response to an abnormal event (Bürki-Cohen, 2010; Green, 1985; Landman et al., 2017a; Landman, Groen, Van Paassen, Bronkhorst, & Mulder, 2017b; Martin, Murray, Bates, & Lee, 2015; Martin, Bates, & Murray, 2010; Martin, Murray, Bates, & Lee, 2016; Rivera et al., 2014). Concordant with this, there is evidence that impaired cognitive processing following startle causes the pilot to forget the proper standard operating procedures and aerodynamic theory applicable to the situation (Bürki-Cohen, 2010). Recent research has explored the effect of startle and surprise on pilot performance. Results show that there are detriments to the pilot's ability and response time to apply standard operating procedures during startling stimuli and unexpected events (Casner et al., 2012; Martin et al., 2016; Schroeder, Burki-Cohen, Shikany, Gingras, & Desrochers, 2014).

Startle and the subsequent cognitive impairment has been documented as a factor in reported pilot malfunction during an emergency (Martin et al., 2016; Rivera et al., 2014). Rivera et al. (2014) analysed incident and accident reports in the Aviation Safety Reporting System (ASRS) from 1994 to 2013. They found 902 reports of surprise and 134 reports of startle. Surprise has a cognitive-emotional response to a stimulus which is similar to startle, however it can also be elicited due to the absence of an expected stimulus (Rivera et al., 2014). The analysed reports only included startle and surprise in reference to flight crew behaviour, excluding flight attendants, passengers or mechanics (Rivera et al., 2014). Of the incidents encoding startle, 37% involved high intensity stimuli which interrupted an ongoing task or elicited a protective reaction (e.g. ducking). All of the 902 surprise incidents were consistent with the definition of surprise. The majority of the incidents included an unexpected event or the absence of an expected event (Rivera et al., 2014).

ASRS reports are voluntary therefore many incidents are likely to go unreported. Thus startle and surprise are likely to be more prevalent in actual operations than the statistics reported by Rivera et al. (2014). Rivera and colleagues' (2014) analysis suggests that startle can be distracting, and interruptive, negatively impacting on the safety of flight deck operations. Green (1985) investigated the relationship between pilot error and aviation crashes and three types of stressors that pilots commonly encounter; environmental, acute reactive, and life events. Reviews of experiments, surveys, and accident and incident reports, showed that acute reactive stress is a factor in many accidents (Green, 1985). Pilots that dealt well with acute stress attributed their performance to simulation training (Green, 1985). It appears imperative to prepare pilots for unexpected, unusual and distracting events to enhance their ability to recover from them.

Pilots that have failed to recognize relevant cues to a developing situation and are surprised by an emergency event occasionally yield to an instinctual reflex to increase their distance to the ground by pulling back on the control column stick (Bürki-Cohen, 2010). The Colgan Air flight 3407 crashed into a house killing 50 people following a stall on an instrument landing approach (NTSB, 2010a). The Captain had a total of 3379 total flight hours and the first officer had 2244 total hours (NTSB, 2010a). A stall on approach is an emergency event that is routinely practiced by pilots; the first officer on Air Colgan flight 3407 had performed between 600-1000 approach-to-stall recoveries (NTSB, 2010a). The proper stall recovery is to lower the nose and apply full power, however when the aircraft warned the captain of the impending stall the captain responded by pulling back on the control column stick, and increasing power to only 75% (NTSB, 2010a). The reaction of the captain of flight 3407 is consistent with an instinctive reflex following startle. Loss of control includes significant, unintended departure of the aircraft from controlled flight, the operational flight envelope, or usual flight attitudes, including ground events (Jacobson, 2010). Loss of control incidents such as the Air Colgan tragedy are one of the leading causes of fatalities in the worldwide commercial jet fleet (Advani, Schroeder, & Burks, 2010; Belcastro & Foster, 2010). Although commercial aviation is overall incredibly safe with very few such events, accidents do sometimes occur. Thus, inappropriate responses due to startle may increase the probability of a negative outcome following an inflight abnormal incident.

Martin et al. (2016) investigated the effects of a startling stimulus on flight performance. They tested pilots on a simulated task involving two hand-flown instrument landing system approaches where the weather was such that a missed approach would be required on reaching the decision altitude. On the first approach, there was a startling stimulus in the form of a cargo fire bell and then an immediate loud bang. Following the first missed approach the pilots were vectored for a second approach. The second approach did not use a startling stimulus; the pilots were required to commence a standard missed approach when they failed to become visual at the minimum altitude required to allow pilots to land on a runway. Height loss during this second missed approach was compared to the height loss of the missed approach following the startle. Martin et al.'s (2016) results indicated that when a startling stimulus immediately preceded a decision-making event, the performance of one-third of the pilots was impaired. The results also indicate that the responses and recovery from startle showed individual variation; some pilots were only slightly affected and recovered quickly, while others were badly affected and took some time to recover. Experience (total number of flying hours and time on aircraft type), and age did not correlate with post-startle performance. It was concluded that the disruption in flying performance and response to abnormal flight events is likely due to startle.

Martin et al.'s (2016) research is one of the very few empirical investigators of the effect of startle on pilot performance in abnormal events. However, there is no evidence presented that shows that the cargo fire bell produced a startle reaction from the pilots. Startle leads to a number of physiological reactions, such as tachycardia, bradycardia, changes in skin conductance, eyeblink, and pupil dilation (Rivera et al., 2014), however none of these were measured in Martin et al.'s (2016) study. It is possible that the cargo fire bell sounding distracted the pilots. Distraction is known to cause impairments in flight performance (Barnes & Monan, 1990). Another possibility is that the cargo fire bell may have confused the pilots as there was no cargo fire to deal with, therefore, the bell was simply an extraneous sound and false alarm, and such confusion could have also led to the impaired performance. Therefore, an improvement on this study would be to have a physiological measure to provide evidence that startle occurred.

Recently, Landman et al. (2017b) investigated the performance of airline pilots during expected versus unexpected stall events. In the unexpected condition pilots were misled and distracted by the experimenter by being asked to focus on irrelevant controls, and asked to give a rating on a sickness scale, before the unexpected event. Likert scales, heartbeat interval durations, and galvanic skin response indicated that pilots were surprised by the unexpected aerodynamic stall. Surprise negatively impacted on their stall recovery, with only three quarters of the pilots successfully adhering to the standard recovery procedure. In the surprise condition mental workload was increased. However, pilots were able to recover without major altitude loss or overspeed indicating that performance was adequate. Although Martin et al. (2015) have suggested the use of distraction to help induce startle, that pilots were highly distracted before the event in Landman et al.'s (2017b) study, will confound the effect of startle and distraction on performance. Therefore it cannot be concluded that startle caused the response impairments, as it would be equally as likely to have been caused by the distraction.

Lack of variation in the presentation of abnormal fight events during training may not be providing pilots with generalizable skills. Casner et al. (2012) investigated the effectiveness of airline pilot training for abnormal events. They sought to establish whether routine simulation training teaches pilots skills that generalise to novel abnormal events. They tested eighteen pilots (nine captains and nine first officers) on three abnormal events: aerodynamic stall, low level wind shear, and engine failure on take-off. These abnormal events were presented in an expected (how they are practiced in airline training) and unexpected fashion (novel situations that could be encountered in real life). When abnormal events were presented in routine ways the pilots reacted appropriately with little variability. However, when they were presented unexpectedly, the responses were less appropriate, more variable, and some pilots were unable to recognize the nature of the event (Casner et al., 2012). The authors suggested that the training system may be faulty because abnormal flight events are presented and practiced in the same manner every time, encouraging pilots to rote learn the singular event signs and responses. Due to this rote learning pilots may not be trained to the level of true expertise as they were unable to generalize their skills to the presentation of events they are well versed in when they were presented in an unexpected manner (Casner et al., 2012). As unexpected events have been shown to lead to startle (Ziperman & Smith, 1975). Casner et al.'s (2012) performance decrement in unexpected events findings may actually be the result of a surprise and/or startle reaction.

Casner et al.'s (2012) study had some obvious methodological flaws. During the briefing the experimenter informed the pilots that their personal skills were not being assessed, however their dependent variables were various performance measures. This briefing may have influenced the pilots not to perform naturally or at their optimum, as they would in a real-life emergency. As well as this Casner et al. (2012) tested the pilots in their normal role (either first officer or captain) then had a confederate pilot in the remaining position. The confederate pilot regardless of cockpit position would only respond or act when requested by the participant pilot. Encountering abnormal events is a situation where pilots are trained to rely on one another and use teamwork to problem solve. Casner et al.'s (2012) experimental paradigm possibly created an unnatural environment for the pilots leading to impaired performance. First officers are subordinate to the Captain, although they can both act as either the pilot-flying or the pilot-not-flying. The Captain would generally take over in an emergency event and has the ultimate responsibility for the welfare of the flight (Nevile, 2001). A subordinate captain would not be very realistic. Therefore, in attempt to keep the first officers in their normal environment by having a confederate pilot as Captain, the researchers may have produced the opposite effect; an abnormal environment. Simulator conditions should have been kept consistent with normal airline testing where pilots are generally tested as a team or individually. A further limitation of this study is that the results of captains and first officers were not separated in statistical analysis so there was no control or comparison to see whether the cockpit positioning affected the results. The proposed study is designed to extend Casner et al.'s (2012) study using tighter controls, solitary general aviation pilots, as well as measuring several physiological parameters that would be indicative of the startle response.

Consistent with Casner et al.'s (2012) conclusions, familiarity due to practice has been shown to impact performance. McKinney Jr and Davis (2003) examined the effects of deliberate practice on pilot decision making during emergency events. They defined a wholly practiced event as a scenario including a malfunction that pilots had deliberately practiced. Whereas a partially practiced event was one where the malfunction occurred within a wider emergency situation which was novel, and had not been practiced. Experienced pilots rated a series of US Air Force Aircraft accident reports. They found that when events had been completely practiced previously performance and decision making was optimal. It was theorized that the practice aided the performance and decision making, as responses were automated and the pilot was able to recognize important cues (McKinney Jr & Davis, 2003). Furthermore, effectiveness in the wholly practiced scenarios was also related to pilots' total number of flight hours, but not in the partially practiced scenarios. However, when the event was only partially practiced (including a novel component that pilots could not have prepared for more experienced pilots made mistakes in the action selection phase and less experienced pilots made mistakes in the evaluation phase. This research suggests that training can improve performance on wholly practiced emergency situations. However this training does not improve pilot's skill for novel elements in emergency events.

When pilots are given a distracting flight task before an abnormal event is induced, the abnormal event appears to induce startle (Landman et al., 2017b; Schroeder et al., 2014). Schroeder (2014) trained pilot participants on two stall manoeuvres; high altitude and low altitude; 73% of pilots applied correct procedure first time for both stall types, but eventually they were all trained to proficiency. The surprise scenario consisted of a global positioning system (GPS) approach. The instructors informed pilots that when they reached the missed approach point (MAP) at 480ft they were going to hand-fly the aircraft to the holding fix where weather may or may not be a factor. However, before they reached the MAP they experienced a large tailwind which rapidly induced a stall at 2100ft. As the pilots had the expectation that there would be an event later in the flying, the early scenario led to startle. It was reported that pilots appeared ‘flustered’, suggesting startle or surprise reaction. Only 22% of the pilots applied the correct stall recovery procedure when they were surprised. It was concluded that unexpected abnormal events led to impairment in pilot’s response and it is important that surprise scenarios are developed and incorporated into training (Schroeder, 2014). However, this research also has the same drawbacks as the Martin et al. (2016) study as there was no physiological measure of startle. As suggested by Bourne Jr and Yaroush (2003) the strength of a stressor can only be determined by measuring physiological and subjective responses of an individual.

Researchers have suggested that pilot training methods should embrace the scenario-based training for upset recovery and invoke some surprise instead of using today’s scripted approach (Advani et al., 2010; Casner et al., 2012; McKinney Jr & Davis, 2003). The FAA has also recommended implementing startling and surprising situations during pilot upset recoveries (FAA, 2015). However, there is a lack of research and technology for methods that induce startle during simulator sessions (Advani, Schroeder, & Burks, 2010). Suggested methods include: creating immersion by making simulator training environment more like an airplane (e.g. wear uniforms, more realistic air traffic communication), increasing workload via the addition of distractions, and invoking a startling situation (e.g. wake upset) after a distraction stimulus (Bürki-Cohen, 2010; Martin et al., 2015; Martin et al., 2016). Martin et al. (2016) suggest that it is possible to program training scenarios in flight simulators that deviate from pilots’ expectations, and thus foster startle or surprise. Airlines and regulatory agencies are becoming increasingly aware of startle and its associated negative effects during abnormal flight events; however, to date there has been little research in an operational context that quantifies the potential effects of startle among airline pilots, and none on general aviation pilots (Martin et al., 2016). Furthermore, the empirical research concerning startle within aviation has lacked physiological measurement.

Understanding factors that affect pilot performance during abnormal flight events is crucial in developing techniques that increase the safety of aviation. To understand and to develop strategies addressing the startle response, it is important to ensure that the simulator conditions designed to produce startle are actually leading to a physiological startle response. Some researchers are currently postulating that startle underlies the degraded performance in unexpected aviation events (Landman et al., 2017a, 2017b; Martin et al., 2015; Martin et al., 2010; Martin et al., 2016; Schroeder et al., 2014). However, most of the empirical research investigating the effect of startle on performance, uses loud sudden noises or distraction (Landman et al., 2017b; Martin et al., 2016). To support the theory that startle due to the unexpectedness of the emergency event is causing the degradation in performance, there needs to be physiological evidence that startle and surprise can occur due to unexpectedness i.e. without a loud noise or distraction.

**Startle and Surprise Conceptual Model**

Recently Landman et al. (2017a) proposed a conceptual model of startle and surprise in terms of sense-making and decision making (Figure 1). The model is an extension of Klein, Phillips, Rall, and Peluso's (2007) data-frame model of sense-making. Cognitive psychology research and theory postulate that knowledge, plans, and theories are grouped in mental structures called schemata or frames (Klein et al., 2007; Landman et al., 2017a). It is thought that information is processed using the currently active frame (Landman et al., 2017a). According to the model, if there is a mismatch between the active frame and the perceived information, a frame switch (reframing) may be required (Landman et al., 2017a). When the active frame and the perceived information are not concordant, the mismatch can lead to surprise or startle (Figure 1). Reframing is a controlled process requiring effort, and reasoning. Thus reframing is susceptible to the negative impact of stress. Stress may lead to the issues with reframing such as choosing an incorrect frame, confusion, or loss of situational awareness (Landman et al., 2017a). The conceptual model is illustrated in Figure 1 which was designed by Landman et al. (2017a).

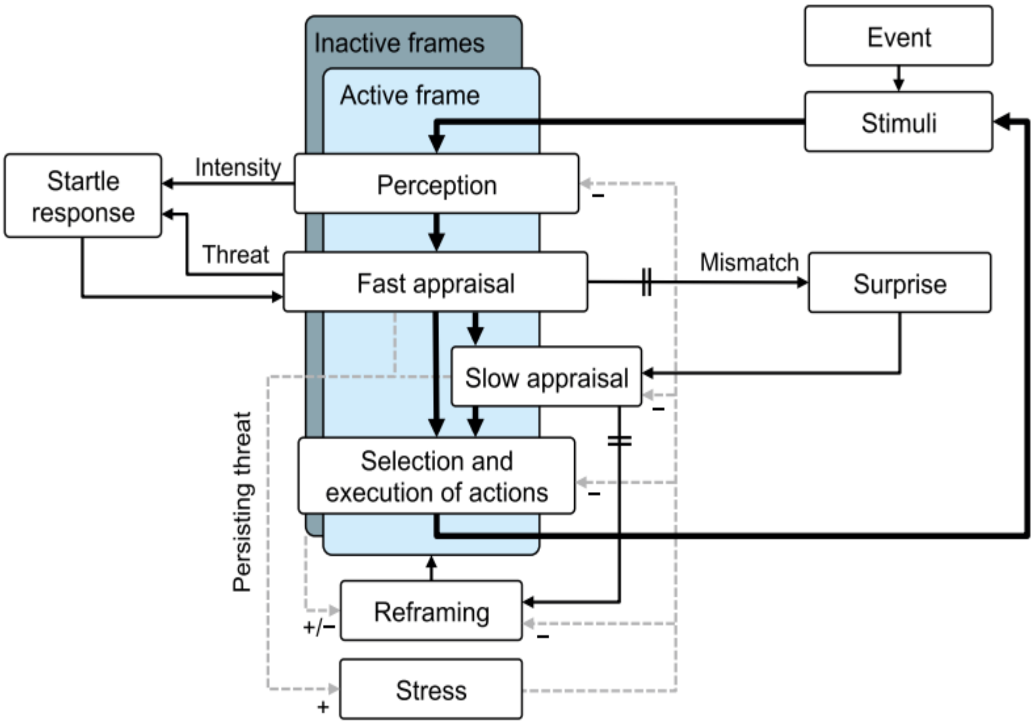



*Figure 1.*Conceptual model of startle and surprise. Solid lines indicate sequenced events. Dashed lines indicate potential influences, with plus signs indicating an increasing effect and minus signs indicating an impaired effect. Double lines indicate thresholds. Reprinted from “Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise” by A Landman, E Groen, MM van Paassen, AW Bronkhorst, and M Mulder, 2017, Human Factors, 59, p. 1163. Copyright 2017, by the Human Factors and Ergonomics Society.

## Expertise and Performance in Abnormal Flight Events

Pilot’s levels of skills and experience can vary greatly. Previous encounters with abnormal flight events may lead a pilot to find an emergency situation challenging, whereas a pilot with no prior experience with the event may find the situation hopeless (Martin et al., 2015). Experts have been found to solve problems faster and consider fewer alternative solutions than do novices (Kirschenbaum, 1992). A study by Li et al. (2001) investigated the factors associated with pilot error in aviation crashes. They examined the prevalence and correlates of pilot error in a large sample of aviation crashes using multiple data files compiled by the NTSB over 14 years. The analysis focused on examining the associations between pilot error and pilot characteristics. It found pilot error to be stated as the cause of 38% of 329 major airline crashes, 74% of 1627 commuter/air taxi crashes, and 85% of general aviation crashes. Additionally the probability of pilot error in general aviation crashes decreased as total flight time and certificate rating of the pilot increased. Thus, expertise may be protective against poor performance in unexpected emergency events.

Research indicates that when a task is well learnt, it becomes more resistant to the negative effects of stress (Baddeley, 1972; Staal, 2004; Wiggins, Stevens, Howard, Henley, & O'Hare, 2002). This is thought to be due to well learnt and well-practiced tasks being committed to long term memory (Staal, 2004; Wickens et al., 1993). Tasks are committed to long term memory through frequent retrieval, activation, rehearsal and recollection (Staal, 2004). When procedural information and knowledge are committed to long term memory, recognition of the cues and retrieval of the related information becomes direct, automatic, and easy (Staal, 2004; Wickens et al., 1993). Automatic retrieval of well-known information from long term memory, is likely to be relatively immune to the negative impact of stress (Wickens et al., 1993). Concordant with this an expert pilot may immediately recognise an emergency or failure, this automatic processing will consume less time and mental resources facilitating the logical reasoning process. (Stokes, Belger, & Zhang, 1990; Wickens et al., 1993).

### Expertise and Startle

Empirical research on expertise and startle in aviation has failed to find any positive effects of increased total flight time on performance in simulated tasks (Casner et al., 2012; Martin et al., 2016). Martin et al. (2016) found expertise measured in terms of total flight hours had no effect on pilot performance when faced with a critical flight decision after startle. Furthermore, Casner et al (2012) also failed to find a protective effect of expertise on performance on unexpected abnormal events. McKinney Jr & Davis's (2003) research found that performance in wholly practiced scenarios was positively related to pilots' total number of flight hours, but not in the partially practiced scenarios. Furthermore, McKinney (1993) found that flight lead pilots with more experience (over 500 hours) showed more degraded diagnostic performance on non-routine malfunctions, in comparison to less experienced (less than 500 hours) flight lead pilots. This suggests that true expertise which would improve performance in emergency flight events may not be revealed via total flight hours. This research suggests that overall experience as measured by flight hours does not necessarily produce expert performance in startling and unexpected events.

### Effect of Expertise and Pilot Type

Previous empirical research concerning abnormal events and startle has investigated airline pilots whereas the proposed study will use general aviation pilots. There are rigorous experience and standard requirements to become an airline pilot. As airline pilots would be at the upper level of experience, more or less flight time might only impact marginally on performance (Li et al., 2001). Airline pilots require a minimum of 1500 hours for certification. Differing amounts of total flight time at higher experience levels (e.g. 1500 hours and upwards) between the participant airline pilots may not translate to any qualitative differences in performance. This may explain the lack of observed effect of experience in some of the studies investigating startle and response to abnormal events (Casner et al., 2012; Martin et al., 2016). General aviation pilots do not go through the same rigorous training and regular assessment, and therefore there is likely to be a larger variation in ability and experience in a population of general aviation pilots (e.g. 50-250 hours). Using general aviation pilots as the participants in a research may be beneficial in uncovering the effects of experience on the ability to respond optimally to an abnormal aviation event.

## Expertise and Information Gathering Systems

Research on expertise has shown qualitative differences between experts and novices in different fields. Kirschenbaum (1992) investigated information gathering strategies used for situational assessment and decision making by twelve Navy officers differing in experience (students, instructors, and former submarine commanding officers). The officers performed a decision task involving response to a passive sonar target. Computer processed and raw sonar data were presented in matrix form where values could only be seen via selecting the box with the mouse cursor. A computer program traced participant's information gathering sequence. Participants were given 90 seconds to examine the matrix of data, and instructed to identify the sonar target and indicate their response to the target. Typical expert information search strategy consisted of smaller total quantity of information used (low number of looks), even divide between raw and processed information (processed data is raw sensor data transformed via algorithms which are easier to interpret e.g. speed and distance), and emphasis on history (integration of information over time) and sets (short transition time between technically related information). The typical novice search strategy consisted of high number of looks (large total quantity of information used) and transitions driven by ease (looking at display items that were close to one another as opposed to technically related). Experts also showed superior situational understanding and decision accuracy. Although Kirschenbaum's (1992) sample size was small, the results provide evidence for qualitative differences in the way experts and novices approach a problem-solving task. Exploring pilot's pattern of information gathering during the encounter of an unfamiliar emergency flight event may provide insight useful for emergency event training.

Wiggins and O'Hare (1995) investigated information processing and search patterns of pilots from different levels of expertise during simulated pre-flight decision making with different time constraints. Wiggins and O'Hare (1995) used a computer-based process tracing methodology to record information acquisition sequence, type of information accessed, decision of whether or not to undertake flight, and response latency. More expert pilots indicated that they would undertake the flight compared to intermediate and novice pilots.

Intermediate pilots accessed a greater number of screens, and seemed to be trying to acquire as much information as possible during the allocated time. This research suggests that expert pilots are better and more efficient at information acquisition and integration (Wiggins & O'Hare, 1995). This finding is concordant with previous research where, O'Hare, Wiggins, Batt, and Morrison (1994) showed that intermediate type pilots are over-represented in aviation accidents caused by decision making errors. This research indicates that compared to pilots of intermediate experience, experts have better decision making skills which allow them to quickly identify a situation and then efficiently implement an appropriate information acquisition strategy.

**Eye Tracking**

Eye tracking is a valuable tool which can be used in addition to normal performance measures. Eye tracking provides data regarding participant's fixation positions, screen fixation durations, and the scan path structure of a presented stimulus (Raschke, Blascheck, & Burch, 2014). This information can be recorded to later study the task solution strategies and cognitive workload from study participants (Raschke et al., 2014). Pupil dilation also measured via eye-tracking has been shown increase during the startle reaction (Bradley et al., 2005; Rivera et al., 2014), with an increase in autonomic arousal (Bradley, Miccoli, Escrig, & Lang, 2008), and with increasing cognitive workload (Bradley et al., 2008; Einhäuser, Stout, Koch, & Carter, 2008; Hyönä, Tommola, & Alaja, 1995; Kahneman, 1973; Marinescu et al., 2018; Marshall, 2002). Eye tracking analysis can also be used to find common eye movement patterns, which can be interpreted as similar cognitive strategies to perform a given task (Raschke et al., 2014). Eye tracking data is commonly analysed in terms of fixations (pauses over informative Areas of Interest (AOI)), and saccades (rapid movements between fixations) (Salvucci & Goldberg, 2000). Visual and cognitive processing is thought to occur during fixations (Salvucci & Goldberg, 2000). Eye tracking data will be collected in the current proposed study to investigate whether there are common information search strategies used by pilots that have superior performance on unexpected abnormal flight events. This analysis may reveal new information which may help to inform training of pilots in preparation for unexpected abnormal flight events.

**The Current Investigation: Research Questions**

1. Do simulated unexpected emergency flight events cause physiological startle?
2. What are the differences in information processing when comparing responses to an expected emergency event and an unexpected emergency event?
3. Does physiological startle impair response performance following unexpected emergency events?
4. Does expertise mediate the startle response during unexpected emergency flight events?
5. Does expertise affect information processing differentially during unexpected emergency flight events?

# Chapter 2: Study 1

## Overview

### Method

The first phase of the investigation was a trial study employing university students. Using university students allows control over participant background and experience in terms of piloting a plane. Furthermore, pilots are both expensive and hard to recruit therefore it was helpful to first complete a trial study to help develop experimenter expertise with the procedure e.g. fixing and problem solving frequent computer bugs and issues that occurred in the experiment set up. The study investigated arousal, information processing, and performance during unexpected flight events compared to expected flight events. In this first proof-of-concept study, university students completed first a flight school PowerPoint (Appendix A), and then nine short simulated flights including five training flights, two unexpected emergency events flights, and two non-event flights (Table 1). The first three training flights involved teaching the students how to fly a Cessna 172SP aircraft in a flight simulator. The second two training flights were designed to teach students how to practice an engine failure and an aerodynamic stall. Two of the remaining flights contained either an unexpected engine failure or an unexpected stall, the other two flights had no events and one was used for baseline data (Table 1). Flight data, eye movements, and heart rate were recorded during these simulator tasks to assess response to abnormal events, information gathering systems, and startle response respectively.

### Hypotheses

It was hypothesized that university students would have higher heart rates and larger pupil dilation during the unexpected flight events (engine failure and stall) compared to the expected flight events and either baseline (heart rates), or before the events (pupil dilation). It was also hypothesized that university students would have a higher heart rate and larger pupil dilation during the expected events compared to the baseline flight or before the events, respectively. It was hypothesized that individuals would spend more time looking at the cockpit flight display screen in the unexpected engine failure as they needed to spend time appraising the situation and they had not been taught to focus on the external environment during engine failures.  It was hypothesized that university students would spend less time looking at the flight displays in the unexpected stall as it is a quick event and they may not realise it is occurring. Furthermore, it was hypothesized participants that crash in the unexpected engine failure would spend more time looking at the flight displays than students that landed safely, as the students that landed safely will concentrate on finding a place to land. It was hypothesized that in the unexpected events university students would have a poorer performance (e.g. more crashes in engine failure, and more altitude lost in the stall) compared to the expected events. It was also hypothesized that participants that had poor performance would show high arousal analogous to the startle reflex when compared to participants that performed well.

## Method

### Participants

Forty first and second year students from the University of Otago were recruited to undergo a flight simulator study. Only twenty one students were able to pass an orientation flight test (described later) and were therefore included in the study. The students were recruited through a psychology database where they obtained course credit for participation in studies run by the psychology department. The ages of the students ranged from 18 to 21 years old, and there were 12 female participants and 9 male participants.

### Apparatus

Students were tested on the commercially available Microsoft Flight Simulator X (MFSX) set in a custom made grey fibreglass cabin with an overhead light (Figure 2 and

Figure 3). The MFSX program was equipped with scenery updates including the ‘Southern lakes adventure: Wanaka and Tekapo’, ‘Real New Zealand Marlborough’ (“Godzone Virtual Flight”, n.d.) as well as the Orbx New Zealand South Island Scenery (Orbx Simulation Systems, 2017). These scenery add-ons were used to enhance the quality and realism of the visual graphics.

*Figure 2.* The flight simulator’s custom made fibreglass cabin. The cabin is facing a blank white wall where the external imagery is projected.

A glass window at the front of the cabin mimicked the windshield of an aircraft by facing a blank white wall background on which the external image was projected via a high-resolution Viewsonic LS820 Full HD 1080p, 0.23 ultra-short-throw projector (Figure 2 and Figure 3). Just below the window, on a desk inside the fibreglass cockpit, flight instruments were presented on one 20” Viewsonic liquid crystal display (LCD) flat panel monitor, below the LCD were the flight controls (Figure 3 and Figure 4). The monitor displayed both the primary flight display (PFD) and the multifunction display screen (MFD) (Figure 5). The PFD displayed the basic flight instruments, such as the airspeed indicator, the altimeter, the attitude indicator, the directional gyro and vertical speed indicator (Figure 5). The MFD presented a moving map display (GPS) (Figure 5). The simulated aircraft was controlled by Cirrus II precision flight controls (Figure 4). These fully functional analogue flight controls are similar to those of a Cessna 172 aircraft featuring all electrical switches, carburettor heat, flaps, gear, and throttle quadrant, along with a metal yoke (Figure 4). Participants were seated facing the window viewing the external image and LCD screen, simulating the feeling of being inside the cockpit of an aircraft (Figure 2 and Figure 4).

*Figure 3.* Flight simulator set up, a photo from inside the cockpit. The image includes the window and projected external image, the flight display (PFD and MFD) displayed on a LCD screen situated below the window, and the Cirrus II precision flight controls situated below the PFD. To the left of the LCD is a PowerLab running the ear mounted plethysmograph.

*Figure 4.* Cirrus II precision flight controls set up located below the LCD screens, used to control the plane.


*Figure 5.* The flight display; the PFD and MFD displayed on the LCD monitor screen inside the cockpit.

The dependent variables measured from each experimental flight were eye-tracking, heart rate in beats per minute (bpm), and flight data (e.g. altitude). Demographics were recorded using a paper questionnaire (Appendix B). Eye tracking data associated with participant focus on the PFD and MFD were recorded and analysed with the Gazepoint GP3 eye-tracking device and software. The Gazepoint GP3 system provided data for each eye at 60Hz. Gazepoint pupil dilation has been recently evaluated and found to be viable and useful in research (Mannaru, Balasingam, Pattipati, Sibley, & Coyne, 2017). Heart rate was tracked using an ADInstruments infrared Plethysmograph (Ear Clip II) connected to a PowerLab version 4/SP, which was in turn connected to a laptop running LabChart7 set up adjacent to the simulator cabin (Figure 6). The Plethysmograph has an infrared photoelectric sensor which recorded changes in pulsatile blood flow in the ear. Flight data were recorded using Burlingame software's Flight Data Recorder version 1.4 (Burlingham Games, 2009). This software uses flight simulator universal inter-process communication to record and analyse data directly from Microsoft Flight Simulator X. Outside and adjacent to the simulator cabin, there was two extra screens, one providing external control of the simulator computer and one which was an extension of the simulator computer screen (Figure 6). The screen set-up outside and adjacent to the simulator cabin allowed for experimenter remote control and monitoring of the MFSX program, the Gazepoint software, and the flight data recorder (Figure 6).

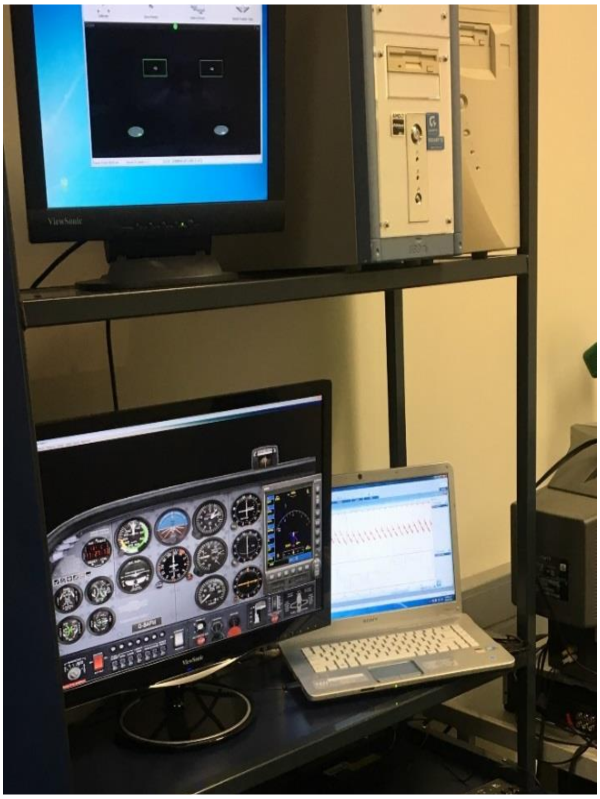


*Figure 6.* Picture of external control screens and laptop. The largest monitor is reflecting the screen inside the cockpit (displaying the PFD and the MFD) and allows for external experimenter control. The laptop runs the LabChart7 connected to the PowerLab and plethysmograph. The small top monitor shows the Gazepoint control screen.

**Procedure**

Each session took between 1 to 2 hours depending on how fast students completed each flight. Each session only involved one student at a time. Students were first asked to read an information sheet about the study (Appendix C), and then asked to fill out a consent form (Appendix D) and a questionnaire regarding their age and gaming experience (Appendix B). Participants were then seated in the flight simulator where they watched a flight school PowerPoint which explained controls in the simulator, the basics of flight, and stalls and engine failures (Appendix A). Participants were then given a headset. Throughout each flight verbal instructions were played through the right side of the headset. The left side of the headset was placed behind the left ear. The headset was configured so that no sound came through the left ear.

Each participant then underwent their first flight which included taking over a mid-air flight at 6000ft over Mount Maunganui (Table 1). This flight was designed to get them used to the feel of the plane and lasted around 5 minutes. No data were collected in this flight. In the next flight participants were introduced again to all of the flight controls, and were then given verbal instruction regarding how to turn on and take off in the plane (Table 1). After participants successfully completed this flight the eye-tracker was set up and calibrated to their eyes, and the plethysmograph clip was attached to the earlobe of the left ear of each participant. The flight data recorder was also turned on and subsequently recorded data for the third flight. The third flight was the orientation flight (Table 1). In this flight participants were required to take off, fly through a series of virtual hoops over Queenstown and circle around to land back at Queenstown airport. Participants were allowed three resets through the orientation flight before their data were excluded from the analysis (orientation test). Nighteen participants out of 40 failed to meet this criteria and were subsequently excluded from the data analysis.

For the orientation flight and the next seven flights the eye-tracking, heart rate, and flight data were recorded for each flight. After the orientation flight participants flew flights four and five (table 1), which were loosely based off normal pilot training and for engine failures and aerodynamic stalls. In these flights students learnt how to respond to an engine failure, and how to induce and respond to a power-off stall. Over all participants, the presentation of these two flights was counter-balanced. For the remaining four flights (flights 6 to 9) the presentation order was randomized and the presentation of the unexpected events was balanced over all participants.

Table 1. *Descriptions of the flights university students underwent in the experiment.*

| ***Flight*** | ***Origin*** | ***Destination*** | ***Event/description*** | ***Approximate flight time (minutes)*** |
|---|---|---|---|---|
| 1 | Mount Maunganui | Not applicable | Participants took over the plane 6000ft high in the air and flew around to get a feel for plane and controls. | 5 |
| 2 | Edward's Air Force base runway (America) | Not applicable | Longest air strip in America (7.5 miles). Participants were re-orientated to the controls via verbal instructions and learnt to take off and fly. | 5 |
| 3 | Queenstown | Queenstown | Orientation flight, participants learnt to taxi, take off, manoeuvre the plane, and land. The majority of the flight consisted of flying through large virtual hoops. If participants needed more than three resets to complete the flight, their data was excluded. | 15 |
| 4 | Pukaki | Omarama | Engine failure training event, participants flew through a series of virtual hoops which led them to safe landing ground where they were informed that the engine will fail. Participants then had to land the plane safely on the flat grassland. | 10 |
| 5 | Dunedin | Taieri | Aerodynamic stall training event. Participants flew through a series of virtual hoops which led them to safe altitude where they were told how to induce a stall and recover. They subsequently completed the stall. | 5 |
| 6 | Invercargill | Ryan's creek | Participants took off into a 40 knot headwind. At 500feet the headwind ceases leading to stall. Due to variation in flight behaviour, only some participants experienced a stall horn. Data was only included from participants that heard the stall warning horn. | 5 |

| | | | | |
|---|---|---|---|---|
| 7 | Glenorchy | Queenstown | Participants took control of the plane mid-flight. The flight is following Lake Wakatipu. An unexpected engine failure occurred after approximately 5 minutes of flight. There were 2-3 safe places to land when the engine failure occured, but the terrain is not optimal. | 8 |
| 8 | Glentanner | Mount cook | No event. Participant flew entire flight eventually landing at Mount Cook, there were a series of virtual hoops on the way which led them to line up with the runway, which they then landed on. | 8 |
| 9 | Te Anau | Not applicable | Participants took over a flight in Te Anau where they were on a ‘rescue mission’ they flew towards indicated lost hikers (using the flight simulators mission compass) and dropped water at the hikers position. | 5 |

# Results

## Data Screening

For each analysis the distribution of the data was examined via boxplots, frequency histograms, and with Shapiro-Wilk tests. The data were considered normally distributed if the Shapiro-Wilk test was not significant ($p>0.05$). Tukey's (1977) method of identifying outliers via boxplots was used for each statistical test. Accordingly, outliers were defined as being more than 1.5 time the interquartile range of the boxplots, and extreme outliers were values that were more than 3 times the interquartile range of the boxplots. In all analyses where outliers were present they were further investigated to determine whether they were plausible values. For example a resting heart rate of 150bpm is highly unlikely, and probably a measurement fault. However, a heart rate of 120bpm is plausible. What is a plausible value was determined from the scientific literature on normative ranges of physiological measurements e.g. Froelicher and Myers (2007). If the value was plausible and the data normally distributed, the outliers were kept. If the data was not normal or the outliers were not plausible the outliers were removed, except for where the sample sizes were small and needed to be conserved. There upon outlier's values were replaced with the next highest or lowest values. If the data were not normally distributed after outliers were addressed, a transformation of the data with the outliers was completed depending on the distribution. A moderately skewed distribution was treated with the square root transformation. A strongly skewed distribution was transformed using a log transformation.

## Heart Rate

For expected and unexpected events, the mean heart rate scores were calculated using LabChart7. For engine failure events, the heart rate scores were calculated for one minute after the event started, for each student, separately. One minute was chosen as following the engine failure pilots took a minimum of 1 minute to land and up to 4 minutes. For the aerodynamic stall events, the mean heart rate for the 30 seconds after the beginning of the event were taken for each student, separately. Thirty seconds were used as the two stall events took between 10 seconds and 60 seconds. For the baseline heart rate, the mean heart rate of the five central (1 minute after start of file until the sixth minute) minutes of the baseline flight which was from Glentanner to Mount Cook (after take-off and before landing), were calculated for each student separately. Heart rate difference between baseline and event heart rate was calculated for each event by subtracting the mean baseline heart rate from the mean heart rate for each event for each participant.

***Engine Failure***

*Mean Heart Rate*

Student participants had a similar heart rate in the unexpected engine failure (*M* = 82.96bpm, *SD* =13.10) and the expected engine failure (*M* = 83.01bpm, *SD* =13.70). During both engine failure events, participants had a slightly higher mean heart rates compared to the baseline flight (*M* = 81.46bpm, *SD* = 12.16) (Figure 7). A one-way repeated measures ANOVA indicated there were no statistically significant differences in participants' mean heart rates during the two different engine failure events and the baseline flight.

*Difference from baseline*

The mean increase in heart rate from baseline in the expected engine failure was smaller (*M* = 1.50bpm, *SD* = 4.92) compare to the unexpected engine failure (*M* = 1.55bpm, *SD* = 5.77). A paired t-test showed that this was not a significant difference.

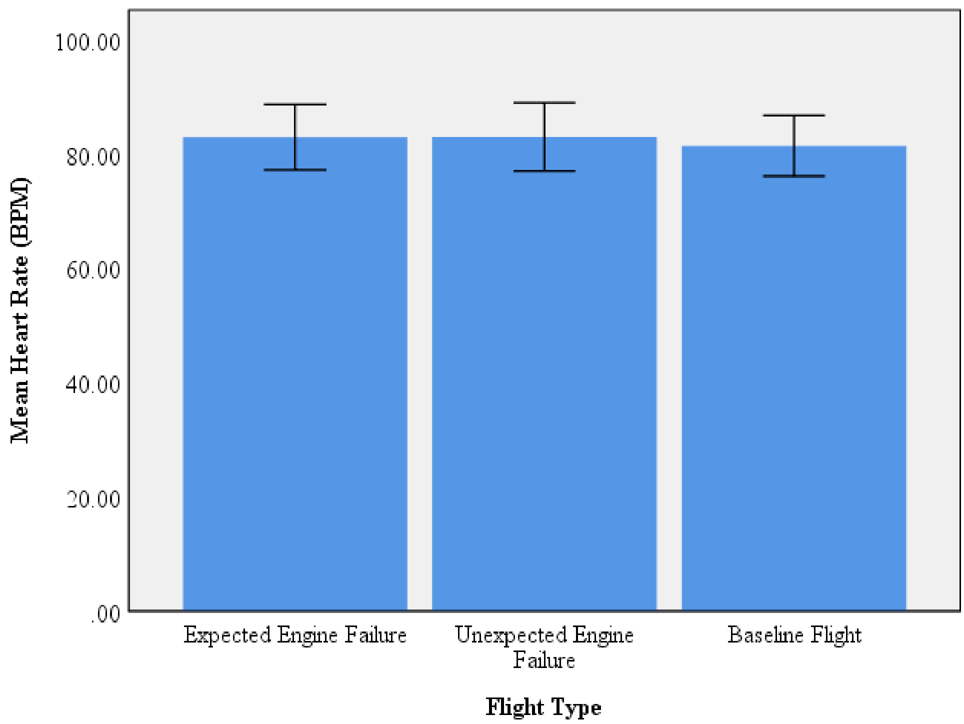



*Figure 7.* Students mean heart rates (BPM) during the expected and unexpected engine failure event, as well as the baseline flight, with error bars (+/- 2 SE).

***Aerodynamic Stall***

*Mean heart rate*

Students had a higher heart rate in the expected stall flight (*M* = 87.68bpm, *SD* =13.69), compared to the unexpected stall flight (*M* = 83.71bpm, *SD* =12.48), and the baseline flight (*M* = 86.06bpm, *SD* = 13.16) (Figure 8). The students mean heart rate in the unexpected stall was also slightly lower than during the baseline flight. A one-way repeated measures ANOVA was conducted to determine whether these were statistically significant differences. There were no statistically significant differences in mean heart rate in participants over the three different flights.

*Difference from baseline*

The mean increase in heart rate from baseline in the expected stall was larger (*M* = 1.62bpm, *SD* = 4.49) compared to the unexpected stall (*M* = -2.36bpm, *SD* = 4.65). The difference scores were normally distributed as shown by a Shapiro-Wilk test ($p > .05$), there was one outlier as assessed by a boxplot. The outlier was a plausible value and due to small sample size the outlier was retained. A paired t-test indicated that this was a statistically significant mean difference of 3.98bpm, 95% CI[0.28, 7.67], $t(8) = 2.484$, $p = .038$, $d = 0.83$. The t-test with the outlier removed was not significant.

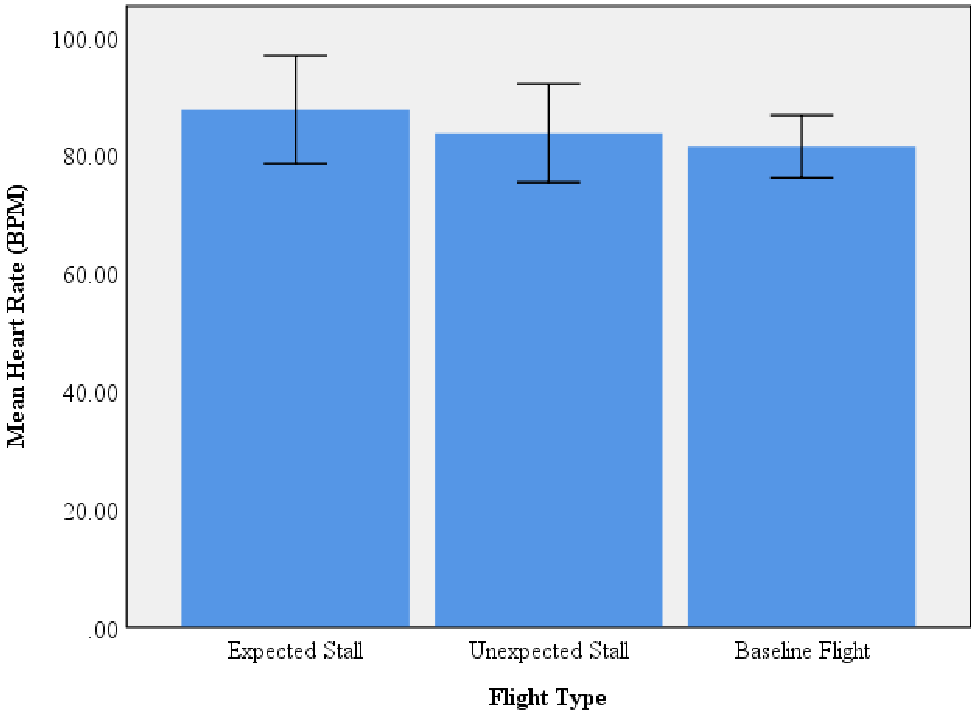



*Figure 8.* Students mean heart rates (BPM) during the expected and unexpected stall events, as well as the baseline flight, and error bars (+/- 2 SE).

**Pupil Dilation**

The Gazepoint software tracks pupil size in pixels (pupil dilation) for both eyes separately over the course of the recording. Using the eye-tracking video it was determined what time the event started and ended. These times were used to calculate separate pupil dilation averages for the left pupil and the right pupil, for before and during each event, for each participant. The averages of the left and right pupil data were then calculated producing before-event and during-event pupil dilation averages for each flight. The difference between these were calculated by subtracting the average pupil dilation before the event from the average pupil dilation during the event. The difference was compared between expected and unexpected flights. Comparing the differences controlled for the variances in the amount of light in each flight as the flights were programmed at different times of the day, for example the Glentanner (baseline) flight was programmed early in the morning therefore the lighting may have affected the pupil dilation.

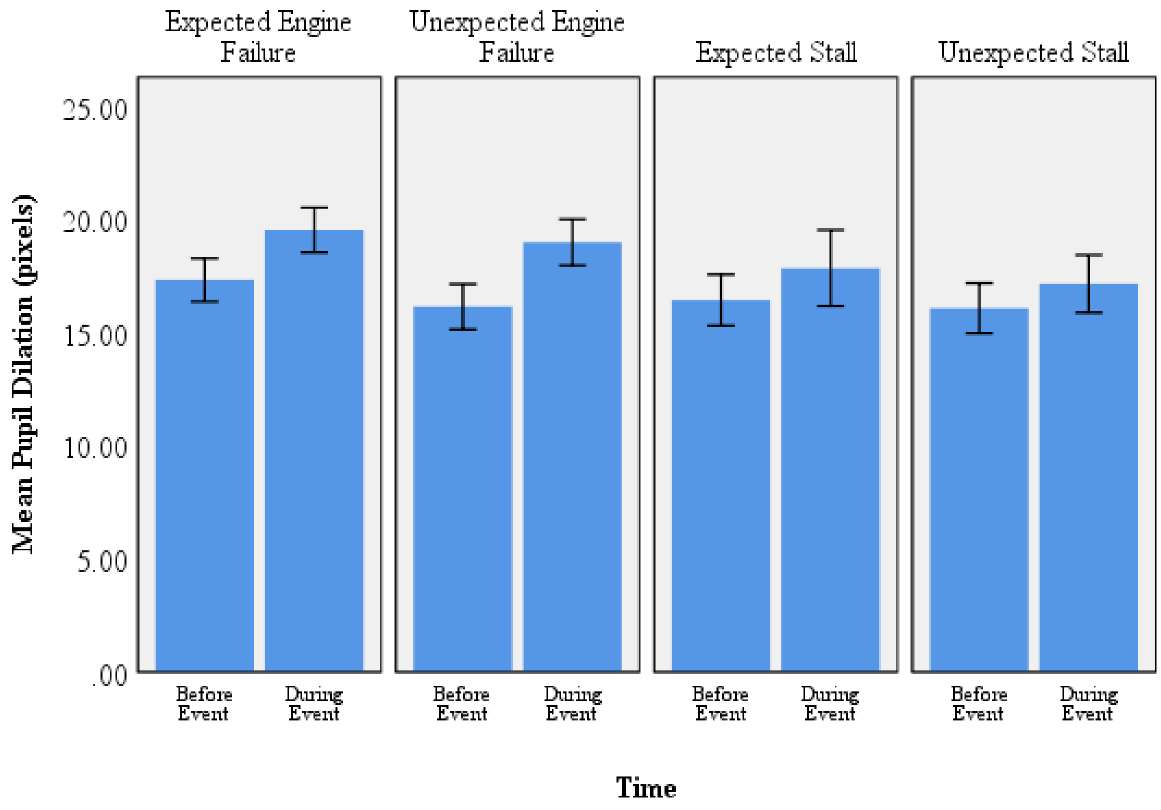



*Figure 9.* Student's mean pupil dilation (pixels) for before and during each expected and unexpected stall and engine failure event, with standard error bars (+/-2SE).

***Engine Failure***

*Expected:* As shown in Figure 9, participants on average had a smaller pupil dilation before the expected engine failure ($M$ = 17.35 pixels, $SD$ = 2.00) and larger pupil dilation during the expected engine failure ($M$ = 19.55 pixels, $SD$ = 2.11). The difference in pupil dilation before the expected engine failure and during the expected engine failure was calculated (pupil dilation before minus pupil dilation during). These difference scores were assessed, no outliers were found and the data were normally distributed as assessed by boxplot and the Shapiro-Wilk test ($p$ > .05), respectively. A paired t-test indicated that this was a statistically significant mean difference of 2.20 pixels, 95% CI[1.863, 2.542], $t$(17) = 13.677, $p$ < .001, $d$ = 3.22.

*Unexpected:* As shown in Figure 9, participants on average had a smaller pupil dilation before the unexpected engine failure (*M* = 16.37 pixels, *SD* = 2.03) and larger pupil dilation during the unexpected engine failure (*M* = 18.92 pixels, *SD* = 2.26). The difference in pupil dilation before the unexpected engine failure and during the unexpected engine failure was calculated. This difference score was assessed, one outlier was found and the data were not normally distributed as assessed by boxplot and the Shapiro-Wilk test ($p < .05$), respectively. The outlier was determined to be not a plausible increase in pupil dilation value and was therefore excluded. Removal of the outlier led to the data having no further outliers and being normally distributed as assessed by boxplot and the Shapiro-Wilk test (p > 0.05), respectively. A paired t-test indicated that this was a statistically significant mean difference of 2.26 pixels, 95% CI[1.940, 3.171], $t(17) = 8.758$, $p < .001$, $d = 2.60$. The t-test with the outlier included was also significant.

*Change in pupil dilation:* The unexpected engine failure led to a larger increase in pupil dilation (*M* = 2.41 pixels, *SD* = 0.68), compared to the expected engine failure (*M* = 2.17, *SD* = 1.12). A paired t-tested indicated that this difference was not significant.

***Aerodynamic Stall***

*Expected:* As shown in Figure 9, participants on average had a smaller pupil dilation before the expected stall (M = 16.47 pixels, SD = 1.70) and larger pupil dilation during the expected stall (M = 17.86 pixels, SD = 2.53). The difference pupil dilation data were assessed and it was found that there was one outlier and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test (p > 0.05), respectively. The outlier was examined and it was determined to be a plausible value and was therefore included. The paired t-test completed with and without the outlier were both found to be significant. A paired t-test indicated that there was a statistically significant mean difference of 1.40 pixels, 95% CI[0.493, 2.298], $t(8) = 3.567$, $p = .007$, $d = 1.19$.

*Unexpected:* As shown in Figure 9, participants on average had a smaller pupil dilation before the unexpected stall (*M* = 16.08 pixels, *SD* = 1.67) and larger pupil dilation during the unexpected stall (*M* = 17.17 pixels, *SD* = 1.91) .The difference pupil dilation data were assessed and it was found that there were no outliers, and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p > 0.05$), respectively. A paired t-test determined that the pupil dilation before and during the unexpected stall had a statistically significant mean difference of 1.08 pixels, 95% CI[0.241, 1.925], $t(8) = 2.965$, $p = .018$, $d = 0.99$.

*Change in pupil dilation:* The expected stall had on average a larger change in pupil dilation (*M* = 1.40 *SD* = 1.17), compared to the unexpected practice stall (*M* = 1.08, *SD* = 1.10). A paired t-test indicated that difference in the change in pupil dilation was not statistically significant.

## Information Gathering and Processing

### *Engine Failure*

*Airspeed Indicator*

Participants spent a lower percentage of their time observing the airspeed indicator in the expected engine failure (*M* = 0.34%, *SD* = 0.36) as opposed to the unexpected engine failure (*M* = 0.63%, *SD* = 0.63) (Figure 10). A paired t-test showed that this difference was not significant.

*Attitude Indicator*

Participants spent a slightly larger percentage of their time observing the attitude indicator in the expected engine failure (*M* = 0.69%, *SD* = 0.80) as opposed to the unexpected engine failure (*M* = 0.61%, *SD* = 0.59) (Figure 10). A paired t-test showed that this difference was not significant.

*Altimeter*

Participants spent a similar percentage of their time observing the altimeter in the expected engine failure (*M* = 0.21%, *SD* = 0.28) as opposed to the unexpected engine failure (*M* = 0.20%, *SD* = 0.28) (Figure 10). A paired t-test indicated that this difference was not significant.

*GPS*

Students spent more time looking at the GPS display during the unexpected engine failure (*M* = 1.04%, *SD* = 0.19), compared to the expected engine failure (*M* = 0.17%, *SD* = 0.21) (Figure 10). The difference data were assessed. It was found that there was one outlier and the data were normally distributed, as assessed by boxplots and the Shapiro-Wilk test ($p > .05$). The outlier was examined and determined to be a plausible value and was therefore retained in the analysis. A paired t-test indication that there was a statistically significant difference of 0.87%, 95% CI[0.475, 1.269], $t(17) = 4.638$, $p < .0005$, $d = 1.09$. When the paired t-test was completed with and without the outlier, the tests remained significant.

*Inside Cockpit (Head Down display)*

Participants spent a lower percentage of their time observing the instrument screen inside the cockpit (displaying the MFD and PFD) in the expected engine failure (*M* = 4.18%, *SD* = 3.11) as opposed to the unexpected engine failure (*M* = 5.08%, *SD* = 3.19) (Figure 11). A paired t-test showed that this difference was not significant.

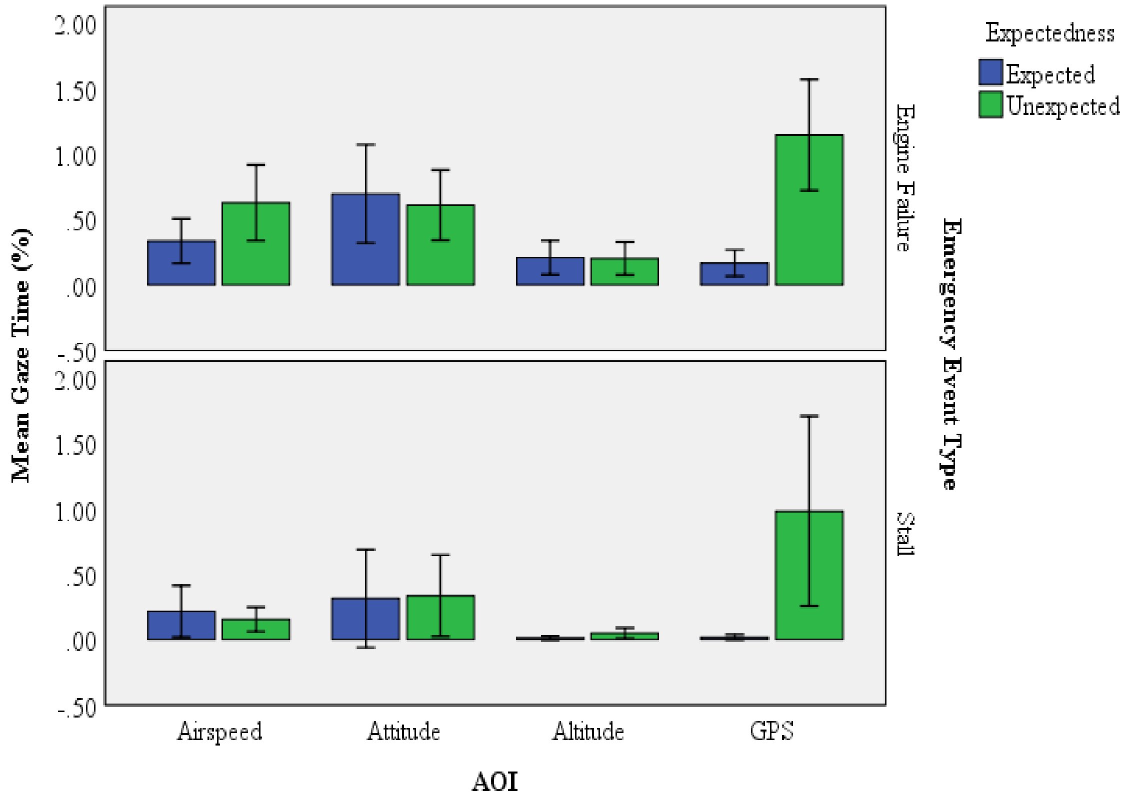



*Figure 10.* Mean percentage gaze time for each flight instrument that participants were taught in the PowerPoint (Appendix A) during both expected and unexpected engine failure and stall flight events, with standard error bars (+/- 2SE).

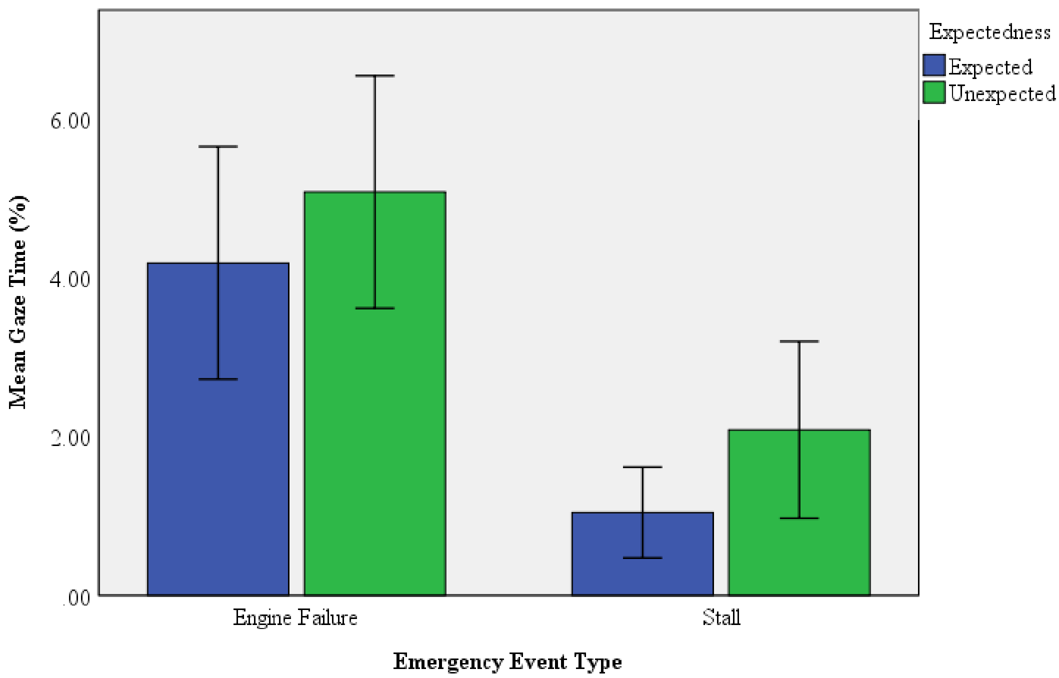



*Figure 11.* Mean percentage gaze time students spent looking inside the cockpit at the PFD and MFD displayed on the LCD screen during both expected and unexpected engine failure and stall flight events, with standard error bars (+/- 2SE).

### ***Stall***

*Airspeed Indicator*

Participants spent a greater percentage of their time observing the airspeed indicator speed indicator in the expected stall (*M* = 0.22%, *SD* = 0.29) as opposed to the unexpected stall (*M* = 0.15%, *SD* = 0.14) (Figure 10). A paired t-test showed that this difference was not significant.

*Attitude Indicator*

Participants spent a slightly lower percentage of their time observing the attitude indicator in the expected stall (*M* = 0.31%, *SD* = 0.56) as opposed to the unexpected stall (*M* = 0.34%, *SD* = 0.47) (Figure 10). A paired t-test showed that this difference was not significant.

*Altimeter*

Participants spent less percentage of their time observing the altimeter in the expected stall (*M* = 0.01%, *SD* = 0.02) as opposed to the unexpected stall (*M* = 0.05%, *SD* = 0.06) (Figure 10). A paired t-test showed that this difference was not significant.

*GPS*

Participants spent a larger percentage of their time observing the GPS in the unexpected stall flight (*M* = 0.98%, *SD* = 1.09) as opposed to the expected stall (*M* = 0.02%, *SD* = 0.03) (Figure 10). The difference data was assessed and it was found that there were no outliers and the data were normally distributed for each flight, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. A paired t-test indicated that this was a statistically significant mean difference of 0.97%, 95% CI[0.368, 1.813], $t(8) = 2.626$, $p = .030$, $d = 0.88$.

*Inside Cockpit (Head Down)*

Participants spent less time observing the flight displays (MFD and PFD) inside the cockpit in the expected stall (*M* = 1.04%, *SD* = 0.86) as opposed to the unexpected stall (*M* = 2.08%, *SD* = 1.67) (Figure 11). A paired t-test showed that this difference was not significant.

## Performance

### *Engine Failure*

During the expected engine failure, one student crashed upon landing. During the unexpected engine failure 14 (66.67%) students crashed the flight and seven (33.33%) students landed safely. The criteria for a crash, was that the students had not successfully touched down on the ground, i.e. there was loss of control on first impact with the ground. Participants that attempted to land in the water were also classed as crashing.

### *Stall*

*Stall Recovery Time*

The time to recover from the expected stall was calculated from the time that the pilot retarded the throttle, and the airspeed was reduced to 60 knots, to when the stall occurred and the plane returned to level flight. The time to recover from the unexpected stall was calculated from the time that the plane reached 500ft till the time that the pilot successfully passed the wind shear and reached 550ft. This method was used as unfortunately the resulting flight video from the trials did not have audio, so the point where the stall horn occurred was not able to be determined. On average students took longer to recover in the unexpected stall (*N = 7*, *M* = 0.20 seconds, *SD* = 0.08) compared to the expected stall flight (*N = 7*, *M* = 0.14 seconds, *SD* = 0.03). Due to the difference in flight configuration this is not comparable (see Table 1).

*Altitude lost*

The altitude lost was calculated as the maximum altitude at the point of the stall minus the minimum altitude reached after the stall. On average students lost a larger amount of altitude in the expected stall flight (*N = 8*, *M* = 922.50 feet, *SD* = 498.16) compared to the unexpected stall flight (*N = 8*, *M* = 80.62, *SD* = 11.65). Due to the difference in flight configuration this is not comparable (see Table 1). Students tended to lose a large amount of altitude in the expected stall as they would pitch the plane up until it was almost vertical, and then after the stall they would pitch it down until it was almost vertical.

*Other*

During the unexpected stall flight, one participant that encountered a successful stall horn, incorrectly pulled back on the throttle.

**Performance and Heart Rate**

***Engine Failure***

Students that landed safely in the unexpected engine failure flight had a slightly lower heart rate (*N* = 7, *M* = 82.37, *SD* = 13.60), than those that crashed in the unexpected engine failure flight (*N* = 14, *M* = 83.33, *SD* = 14.25). An independent t-test showed that this difference was not significant.

***Stall***

There was no correlation between altitude lost, recovery time and heart rate in both the expected and unexpected versions of the stall.

**Performance and Pupil Dilation.**

***Engine Failure***

Student participants who crashed had a slightly smaller pupil dilation (*N* = 7, *M* = 18.99, *SD* = 2.37) compared to students whom landed safely (*N* = 14, *M* = 19.09, *SD* = 2.14), however participants that crashed had a larger increase in pupil dilation after the engine failure (*N* = 7, *M* = 3.16, *SD* = 2.07), compared to participants who landed safely (*N* = 14, *M* = 2.20, *SD* = 0.60). Independent t-tests showed that these differences were not significant.

### ***Stall***

There was no correlation between altitude lost, recovery time and pupil dilation or change in dilation in both the expected and unexpected versions of the stall.

## **Performance and AOI**

### ***Engine Failure***

#### *Airspeed*

Students who crashed in the unexpected engine failure flight spent a larger percentage of their time looking at the airspeed indicator (*N* =14, *M* = 0.86, *SD* = 0.64), compared to students who landed safely (*N* = 7, *M* = 0.14, *SD* = 0.12). The distribution of the data was examined. There were no outliers in both groups, and the data were normally distributed as assessed by boxplots, and Shapiro-Wilk tests ($p > 0.05$), respectively. Due to a violation of the homogeneity of variance assumption. A Welch t-test was used to assess whether these differences were significant. The t-test indicated that there was a statistically significant difference of 0.72%, 95% CI[0.32 to 1.12], $t(13.79) = 3.887$, $p = .002$.

#### *Attitude Indicator*

Students that crashed during the unexpected engine failure spent slightly less time looking at the attitude indicator (*N* = 14, *M* = 0.60, *SD* = 0.55), than the students that landed safely (*N* = 7, *M* = 0.64, *SD* = 0.71). An independent samples t-test showed that this difference was not significant.

#### *Altimeter*

Students that crashed during the unexpected engine failure spent a greater percentage of their time looking at the altimeter (*N* = 14, *M* = 0.22, *SD* = 0.31), than the students that landed safely (*N* = 7, *M* = 0.16, *SD* = 0.21). An independent samples t-test showed that this difference was not significant.

*GPS*

Students that crashed during the unexpected engine failure spent a larger percentage of their time looking at the GPS (*N* = 14, *M* = 1.34, *SD* = 0.1.06), than the students that landed safely (*N* = 7, *M* = 0.74, *SD* = 0.27). An independent samples t-test showed that this difference was not significant.

*Inside Cockpit (Head Down)*

Students that landed safely after the unexpected engine failure spent a lower percentage of their time looking inside the cockpit at the flight display (PFD and MFD), (*N* = 7, *M* = 3.26, *SD* = 1.65), compared to student who crashed (*N* = 14, *M* = 5.92, *SD* = 3.42). An independent t-test revealed that this was not a significant difference.

***Stall***

There were no significant correlations between the percentage time students spent looking at the airspeed indicator, the attitude indicator, the altimeter, the GPS, or inside the screen and stall recovery time, or altitude lost in either the expected or the unexpected stall flight.

## Discussion

### Physiological Measures

It was proposed that participants would show more physiological arousal during unexpected emergency events than in the expected emergency events. Specifically, it was hypothesized that university students would have higher heart rates and larger pupil dilation during the unexpected flight events (engine failure and stall) compared to the expected flight events and either baseline (heart rates), or before the events (pupil dilation). It was also hypothesized that the expected versions of events would show more arousal than the baseline flight.

Consistent with the hypotheses during both expected and unexpected engine failure and stall events there was a significant increase in pupil dilation compared to before the event. These increases in pupil dilation indicate autonomic arousal during both the expected and unexpected events. Contrary to the hypotheses the unexpected stall had a decrease in heart rate compared to baseline, and the expected stall had an increase in heart rate compared to baseline. Furthermore the expected stall event had a significantly larger mean difference in heart rate when compared to baseline in comparison to the unexpected stall. Contrary to the hypotheses, there were no significant differences in heart rate or change in heart rate during the unexpected engine failure compared to the expected engine failure and baseline. Inconsistent with the hypotheses there were also no significant differences in heart rate during the unexpected stall compared to the expected stall event and the baseline flight. The heart rate data do not indicate any differences in autonomic arousal across the flights. However, the significant increases in pupil dilation during both aerodynamic stall events and both engine failure events suggests the presence of a physiological autonomic response.

Contrary to hypotheses the data do not indicate any significant differences in participant's physiological reactions to the expected engine failure compared to the unexpected engine failure. Also inconsistent with the hypothesis there is some evidence that the expected stall led to more autonomic arousal than the unexpected stall, which in turn elicited less physiological arousal than the baseline flight. The expected engine failure and stall, and the unexpected engine failure led to small increases in heart rate of approximately 2bpm when compared to baseline. However, these increases were not statistically different from baseline.

Research that has investigated the effect of the startle response on heart rate has found a mean increase in heart rate of approximately around 7.50-15bpm in response to a loud noise (Chou, Marca, Steptoe, & Brewin, 2014; Deuter et al., 2012; Holand, Girard, Laude, Meyer-Bisch, & Elghozi, 1999). Research investigating high workload has shown a mean increase in heart rate from baseline to high workload condition of approximately 3-6bpm (Fallahi, Motamedzade, Heidarimoghadam, Soltanian, & Miyake, 2016; Grassmann, Vlemincx, Von Leupoldt, & Van Den Bergh, 2017). As well as this, an increase in pupil dilation is indicative of a higher cognitive workload (Bradley et al., 2008; Hyönä et al., 1995; Kahneman, 1973; Marinescu et al., 2018; Marshall, 2002), and autonomic arousal (Bradley et al., 2008). There were also significant increases in pupil dilation from before the flight events to after which suggests increase in workload. Furthermore, the increase in heart rate from baseline for both engine failure event flights, and the expected stall event flight was approximately 2bpm. Although these increases were not significant, it can be speculated that they might indicate a slight increase in workload, in line with the pupil dilation data. Consequently, the data indicate that in all flight events students did not have a startle reaction. Instead the data suggest physiological changes may be indicative of a slight increase in workload during the expected stall and engine failure, and during the unexpected engine failure.

It is unlikely that students were motivated to perform well, which may have impacted the effect of unexpectedness emergency events on arousal. University students enrolled in psychology courses are rewarded course credit for participation (approximately 0.667 credits per hour). These are the students that were recruited for this experiment. Most of the students therefore would not have had a real interest in piloting aircraft and are likely to have just wanted to get through the experiment as fast as possible. Lack of motivation or care about performance may have impacted the probability of startle or stress during the unexpected events. Supporting this idea, Lazarus et al., (1952) suggested that one of the largest problems of the failure-stress technique is controlling for the subject's motivation. For a participant to be stressed, they must have the motivation to do well or avoid failure (Lazarus et al., 1952; Skinner & Brewer, 2002). Thus the effect of stress is generally dependent on the expectation an individual has on themselves (Lazarus et al., 1952; Skinner & Brewer, 2002). Concordant with this, research investigating arousal in response to psychological stressors in terms of motivation level has found a positive relationship between motivation and stress response (Bergman & Magnusson, 1979; Vogel, Raymond, & Lazarus, 1959). For example Vogel et al. (1959) found that high school boys low in achievement motivation but high in affiliation motivation elicited more physiological arousal when tested on a measure of warmth and friendliness compared to a test of achievement ability. While boys with the opposite motivations had higher arousal on the test of achievement ability. Therefore this research by Vogel et al. (1959) found that physiological arousal levels are higher when a more valued goal is threatened compared to when a less valued goal. Lack of motivation from student participants would lead to a low probability of a startle reaction, and explain the small, non-significant changes in heart rate. The study tried to ensure participants were motivated by having a reasonably hard orientation test, where participants needed to be engaged or somewhat skilled to pass. However, due to the length of the study and a minimal baseline level of interest in aviation, motivation likely impacted the degree of physiological responses to the unexpected events from the university students.

This experiment was the students first time flying a flight simulator, and students had very minimal exposure to piloting aircraft before participating (one participant had flown in a small aircraft with a family member once before). Therefore, before practicing the expected flight events, students had little experience (three learning flights (Table 1)). By the time the participants encountered the unexpected flight they would have at least 'flown' five short simulated flights. This may have led to flying in the unexpected events requiring less cognitive workload or being less stressful than the first few flights including the expected versions of the emergency events. This may explain why the expected flight events had a higher heart rate compared to the unexpected flight events. Unfortunately, it wasn't possible to control for this potential order effect as it is very difficult to rely on students that earn course credit to come in for two separate sessions. For example, over 20 participants failed to turn up for their scheduled session. Therefore, all flights needed to be completed in one session, and obviously students needed to be taught how to respond to abnormal flight events before experiencing the events unexpectedly. If the students had a few more training sessions before the expected and unexpected events there may have been a more pronounced effect of unexpectedness. The ideal novice experiment would be to have the students practice flying and emergency events for around five hours before encountering the unexpected versions of events. However, this was not feasible in this trial experiment.

Student participants in this experiment had relatively high baseline heart rates compared to data from other studies and for their age. Other research has consistently reported a mean resting heart rate over a sample of around 70-78bpm (Basner, 2009; Chou et al., 2014; Fallahi et al., 2016; Grassmann et al., 2017; Main, Wolkow, & Chambers, 2017; Orr, Solomon, Peri, Pitman, & Shalev, 1997). Agelink et al. (2001) reported that young male adults (17-25years) have a mean heart rate of 68.9, and young female adults had a mean heart rate of 76.7 (17-25years). Therefore it would be expected that university students ranging from 18-21 years old would have mean baseline heart rate of around 70-75bpm (Agelink et al., 2001). Fallahi et al. (2016) investigated the changes in physiological measures (heart rate and electroencephalography) over different mental workloads for operators in cement, city traffic and power plant control centres. Fallahi et al. (2016) reported a mean resting heart rate of 75.9bpm, which increased with increasing mental workload; high mental workload condition had a mean heart rate of 79.7bpm. Grassmann et al. (2017) investigated cardiorespiratory measures over different mental workloads and found a mean baseline rate of 71.32bpm and in the multiple task (high workload) condition mean heart rate was 77.37bpm. Therefore, in comparison to previous research, the university students had a high baseline heart rates this may have impacted the probability of finding a further increase in heart rate in response to the flight events.

The high baseline heart rate of the university students in the current study could indicate that all flights, even after five training flights, were high workload tasks. Chou et al. (2014) found that participants with a high baseline heart rate had a low startle response. Concordant with this, previous research has also found that concurrent working memory tasks can disrupt the emotional regulation of the startle reflex (King & Schaefer, 2011). The workload may have been already high in all of the student flights indicated by high heart rates. Therefore in the present study high workload and high heart rate during the experimental flights may have moderated any startle response to the unexpected events. In other words, the high baseline heart rate may have impaired the ability to find significant changes in heart rate for the flight events.

It is likely that student participants are not conscious of the danger a real aerodynamic stall at 500ft could engender. This could also lead to small or no change in heart rate in response to the unexpected stall. However, this does not explain a decrease in heart rate when compared to baseline in the unexpected stall. In the baseline flight the students were required to fly through virtual hoops to help them find and land at the airstrip (Table 1). Virtual hoops were used in most of the university student flights as it took away the requirement of navigation. There were no such hoops in the unexpected stall flight. Flying through the virtual hoops in the expected stall may have been harder requiring more cognitive workload than encountering the plane upset and stall horn in the unexpected stall event. High cognitive workload leads to higher autonomic arousal (Hyönä et al., 1995; Marinescu et al., 2018;

Marshall, 2002). Therefore the virtual hoops may be why there was higher heart rates and larger pupil dilation the expected stall flight than in the unexpected stall.

**Information Processing**

It was hypothesized that individuals would spend more time looking inside the cockpit at the MFD and PFD head-down flight displays during the unexpected engine failure as opposed to the expected engine failure. This was anticipated as the participants would have needed to spend more time appraising the situation in comparison to the expected engine failure in which they were briefed on the upcoming event. As well as this, the university students had not been taught to focus on the outside environment to help find a landing spot, and to help safely aviate during an engine failure. It was hypothesized that university students would spend less time looking at the different flight instruments in the unexpected stall as it is a quick event and they may not realise it is occurring. Contrary to both of the hypotheses there were no significant differences in time spent looking at the combined MFD and PFD displays during either of the unexpected events.

The following discussion in the information processing section contains speculation on possible differences in gaze time for the entire MFD and PFD, and the airspeed indicator during the expected and unexpected events and therefore conclusions should be viewed as tentative.

The only significant difference in terms of gaze behaviour between the expected and unexpected versions of both event types, was percentage time spent viewing the GPS. University students spent significantly more time viewing the GPS in the unexpected engine failure and stall compared to the expected engine failure and stall, respectively. This could be because to navigate in the expected events students followed mission compass and virtual hoops, however for navigation in the unexpected events the GPS was used. The participants spent a similar amount of time viewing the attitude indicator, and the altimeter during all stall and engine failure events. There is a possibility that there is an undetected true difference in percentage time spent viewing the airspeed indicator. Although non-significant, participants spent more time looking at the airspeed indicator in the unexpected engine failure and stall compared to the expected engine failure and stall, respectively. However, this is just speculation.

Visual and cognitive processing is thought to occur during fixations (Salvucci & Goldberg, 2000). In this study, fixations are the percentage time participants spent looking at the individual AOIs. The GPS and the airspeed indicator are arguably the most recognisable and familiar displays to non-pilots. The altimeter requires a mental translation, similar to that of reading an analogue clock, and thus requires a degree of effort and learning. The attitude indicator is possibly the most helpful instrument for novices, to ensure stable plane handling as well as enabling the participant to maintain altitude without stalling while gliding during engine failure recovery. The attitude indicator, although visual is not comparable to anything laypeople would see or use in everyday life. The airspeed indicator, is very straightforward and is similar to a car speedometer. Furthermore, students were taught to use the airspeed indicator at each take-off and landing throughout the experiment. The GPS is also similar to any GPS used on the ground e.g. Google maps. Therefore participants would have likely had a degree of familiarity with the GPS. Kirschenbaum (1992) found that novice operator's information gathering strategy was driven by ease, as participants tended to look at controls that were close to each other. Consistent with this, in the unexpected flight events university students appear to be focusing their attention on the controls that are familiar, easy to use, and easy to comprehend (the airspeed indicator and the GPS). This suggests that during the unexpected engine failure university students may have adopted an information processing strategy that reduced their cognitive load.

The unexpected stall was a very quick event, and it is possible that the students did not actually realise that it was occurring. It was common for participants to induce the stall horn via bad plane handling in the training phases, which may have caused the stall horn to be less salient as an indicator of an emergency event. The lack of change in heart rate supports this, however the increase in pupil diameter does not. Participants may have a higher workload due to the plane upset, occurring in the unexpected stall regardless if they did not recognize the event, this would explain the increase in pupil dilation. Participant's pattern of information search in the unexpected stall also supported the idea that they did not recognize that there was an abnormal event occurring. In the unexpected stall the airspeed indicator would be the most useful instrument in the diagnoses of the situation as observing it would reveal that at 500feet there was a sharp drop in airspeed leading to stall. Although it was not significant, University students spent less time viewing the airspeed indicator in the unexpected stall than in the expected stall. If students were trying to diagnose the situation they should have noticed the sharp change in airspeed occurring at the same time as turbulence and loss of smooth controlled flight. The finding that students spent significantly more time viewing the GPS in the stall suggest that students did not recognise the event was occurring, and therefore continued to fly and navigate towards their end goal of Stewart Island. Participants continued to view the GPS, even though it would provide minimal help for recovery in a stall.

Although the differences were not significant, the data indicate that students may have spent longer on average looking inside the cockpit (heads down) at the MFD and the PFD during the unexpected compared to the expected engine failure and stall events (Figure 11). In the expected event flight participants were informed of what was going to occur at the start of each flight. Therefore participants did not need to spend time diagnosing the situation when it occurred. However, during the unexpected event flights, participants would have needed to assess and try and diagnose the situation. This may explain why participants spent longer looking inside the cockpit at the head down display showing the MFD and the PFD during the unexpected engine failure and stall events.

**Performance**

It was hypothesized participants that crashed in the unexpected engine failure would spend more time looking at the displays than students that landed safely. This is because the students that landed safely would have concentrated on the external environment to find a safe place to land. Contrary to the hypothesis there were no significant differences in the amount of time participants spent looking inside at the cockpit (head down display) at the PFD and PFD between those that crashed and those that landed safely. Those that crashed spent significantly longer looking at the airspeed indicator. Participants that crashed spent similar amounts of time looking at the attitude indicator, and the altimeter. Although not significant participants that crashed had a larger mean percentage time looking at the GPS compared to the students that landed safely. As well as this, participants that crashed spent longer overall looking inside at the cockpit (head down display) at the PFD and PFD than those that landed safely, however these differences were not significant. Therefore the tendency for students to spend more time looking at the airspeed indicator and the GPS in the unexpected compared to the expected engine failure is likely to have been a change in information processing strategy that was detrimental to performance.

It was hypothesized that in the unexpected events university students would have a poorer performance (e.g. more crashes in engine failure, and more altitude lost in the stall) compared to the expected events. As hypothesized more participants crashed during the unexpected engine failure than the expected engine failure. However, this may simply be due to the complexity of the landing. Consistent with the hypotheses individuals that crashed in the engine failure had higher heart rates, and slightly larger increases in pupil dilation during the engine failure than participants who landed safely, however these differences were not significant. Heart rate and pupil dilation was not associated with recovery time or altitude lost in the either stall event.

**Conclusion**

Although this study was good for a proof-of-concept, a majority of the physiological and performance measurements were not significantly different between the unexpected and the expected events as expected. Although students had only just learned to fly it was still expected that they may have an autonomic response to unexpected emergency events. Thiese lack of differences were hypothesized to be due to a lack of motivation, existing high workload, or a lack of recognition of emergency cues and danger. The results of this first study indicate that students have higher arousal during both expected and unexpected flight events compared to before the events. This is suggested by and the significantly larger pupil dilations during the event compared to before the event. However, the small increases in heart rate in the unexpected events when comparing to baseline indicate that there was no startle reaction. Instead it is speculated that the small increases in heart rate indicate increases in cognitive workload.

In the unexpected abnromal participants tended to focus on the GPS which provides navigation information as well as an estimated time until flight completion. Therefore the information participants gathered during the unexpected events does not indicate any inclination of trying to assess and safely recover from the events. This may indicate that participants did not recognize the events, or did not care that they had occurred, which also may explain the lack of the hypothesized arousal. The information search strategy could also indicate that when encountering abnormal events the university students relied on the information that was the easiest to process even when that information was not particularly helpful concerning the current flight situation.

# Chapter 3: Study 2

## Overview

*Method*

The second phase of the study investigated the same research questions as Study 1 but with general aviation pilots. The study investigated physiological arousal, information processing, and performance during unexpected flight events compared with expected flight events. General aviation pilots completed a series of seven short simulator flights (Table 2). The first flight was an orientation flight where pilots first flew in the simulator and the experimenter checked all recordings. Amongst the flights were expected and unexpected presentations of two abnormal events; aerodynamic stalls, and engine failures. The expected abnormal event was presented to the participants consistent with normal flight training (CAA, 2012), which pilots would have practiced before. The unexpected abnormal event was presented in a novel and unexpected manner consistent with a real life emergency (Table 2). The other two flights were non-event flights. One of these flights were used for the baseline data (Table 2). Flight data, eye movements, and heart rate was recorded during these simulator tasks to assess physiological responses, information gathering systems, and performance to the flight events.

*Hypotheses*

It was hypothesized that pilots would have a higher heart rate and larger pupil dilation during the unexpected flight events (engine failure and stall) compared to the expected flight events, and baseline (heart rate), or before the event (pupil dilation). It was also hypothesized that pilots would have a higher heart rate and larger pupil dilation during the expected events compared to the baseline flight or before the events, respectively. In unexpected flight events pilots would have needed to assess and diagnose the situation, compared with the expected events where they were informed of what was happening and therefore could skip the assessment phase. Therefore, it was hypothesized that pilots will spend more time looking inside at the cockpit (head down display) during unexpected events. It was hypothesized that in the unexpected events pilots would perform more poorly than in the expected events (e.g. more crashes in engine failure, more altitude lost in the stall, and more incorrect responses) compared to the expected events. It was also hypothesized that those pilots that showed impaired performance would also show high arousal indicative of the startle reflex. Furthermore, it was hypothesized pilots that crashed in the unexpected engine failure would have spent more time looking at the controls than pilots that landed safely, as the pilots may have forgotten to prioritise aviating and focus on the flight displays and instruments. Lastly it was hypothesized that pilots that performed better on unexpected abnormal events would have similar information gathering strategies as revealed by the eye tracking data.

## Method

### Participants

Twenty-two pilots from the local Otago area were recruited through posters at local institutions, word of mouth, and social media advertising. Pilots were given $40 for their participation in the research. Pilots were required to have at minimum solo flight certification. The ages of the pilots ranged from 16 to 61, and their level of experience ranged from 15 hours to 2050 hours. There was 1 female pilot and 21 male pilots.

### Apparatus

The apparatus used was similar to Study 1. The flight simulator set-up remained the same. Pilot participants were not shown the flight school PowerPoint, or given any training flights, as ability was assumed. The information sheet (Appendix E), and the consent form (Appendix F) were altered to reflect the pilot participants and the associated monetary reimbursements for their time. As well as this, demographics were recorded using an online SurveyMonkey questionnaire https://www.surveymonkey.com/r/YYYNR83 (Appendix G).

Furthermore, before each flight pilots received introductory pages describing a fictitious purpose behind each flight, including the current position of the aircraft, the flight's weather forecast, an aeronautical map of the area, and the goal of the flight (Appendix H). The introduction also included the aerodrome charts for the airstrips or airports involved in the flight. The aerodrome charts were sourced from the Aeronautical Information Publication New Zealand website (AIPNZ, 2002).

**Procedure**

*Introduction to the simulator*

Each session took between 1 to 2 hours depending on how quickly pilots completed each flight. Each session only involved one pilot at a time. Pilots first read the information sheet (Appendix E), signed the consent form (Appendix F), and then completed the Survey Monkey questionnaire in an adjacent room (Appendix G). Pilots were then verbally introduced to the simulator and the MFSX program, and briefed on the outline of the general experimental procedure and what was required of them. Afterwards, pilots were seated in the simulator, where the seat position and/or the external image were repositioned to ensure they had optimal viewing of the simulated image through the windshield of the simulator cabin. The plethysmograph was attached to the pilot's ear and the eye tracking camera was positioned and calibrated.

To prepare the pilot for each flight they were handed introductory pages describing a fictitious purpose behind each flight (Appendix H). Before each flight pilots were instructed to read these pages, and after each flight the experimenter set up the next flight while the pilot viewed the charts, maps, and instructions for their next flight.

*Flights*

After viewing the flight briefing information for the orientation flight between Cape Foulwind and Westport, pilots were asked to inform the experimenter when they were ready to start the flight. The experimenter then un-paused the loaded simulated flight and then the pilot completed the task. This process was repeated for all remaining flights which were presented in a randomized order. The randomization was done using https://www.randomizer.org/ (Urbaniak & Plous, 1997).

The aircraft chosen for the experiment was the Cessna 172SP which is the most widely available rental aircraft and is used by most flight schools. This ensured that most of the pilots had some level of familiarity with the aircraft or at least experience with a very similar aircraft. After the orientation flight from Port Foulwind to Westport pilots completed seven further flight simulation tasks which are described in Table 2. Four of these flights included inducing or responding to two types of abnormal flight events; aerodynamic stall (unexpected and expected versions), engine failure (unexpected and expected versions). In the 'expected' simulation conditions the stall and engine failure flight events were presented in a manner that the pilots are likely to be familiar with. These expected abnormal events refer to the standardized practice presentations that are commonly taught in all flight training schools (FAA, 1988, 1999; CAA, 2012). In contrast, in the 'unexpected' flight event simulation condition tasks included stall and engine failure flight events presented in a novel manner with no forewarning. The abnormal events were designed to be similar to how they could present in a real-life situation.

Table 2. *Descriptions of the flights pilot participants underwent in the experiment.*

| *Flight* | *Origin* | *Destination* | *Event/Description* | *Approximate flight time (minutes)* |
|---|---|---|---|---|
| 1 | Cape Foulwind | Westport | Orientation flight. Pilots took off on a small rural air strip and flew for a few minutes, before landing at Westport. This flight was designed to allow pilots to get used to the feel of the simulator | 5 |
| 2 | Invercargill | Ryan's Creek | Pilots were presented with a stall at 500 feet while climbing out after take-off from Invercargill airport in high winds. This stall was created by rapidly shifting prevailing winds from a strong headwind to a slight tailwind. This wind change caused a sudden loss in airspeed leading to a stall horn warning. The correct response was the same as flight 5 (the expected stall flight). | 5 |
| 3 | Pukaki | Omarama | Expected engine failure (informed on briefing sheets). Five minutes after taking off from Pukaki airstrip a total engine failure occured. Pilots were required to find a safe place to land. | 5 |
| 4 | Glenorchy | Queenstown | After taking off from Glenorchy pilots fly along Lake Wakatipu. The unexpected engine failure, occured approximately eight minutes into the flight after the pilots reached the large bend in the Lake and turned towards Queenstown. Pilots were required to find a safe place to land. | 13 |
| 5 | Dunedin | Taieri | In this flight pilots were required to take off from Dunedin airport and fly to a safe altitude and practice a power-off stall. This is where the pilot retards the throttle and pitches the plane slightly up leading to a decay of airspeed. As soon as the stall horn occurs the pilot pitches the plane downwards and moves the throttle to maximum thrust. | 10 |
| 6 | Manapouri | Te Anau | Pilots took off in good conditions and land at Te Anau. There were no events in this flight. | 5 |

| | | | | |
|---|---|---|---|---|
| 7 | Glentanner | Mount Cook | Pilot took off from Glentanner and landed at Mount Cook. There were no events in this flight. The recordings from this flight were used as a baseline measurement. | 8 |
| 8 | Roxburgh | Alexandra | Pilots took off from Roxburgh and landed at Alexandra. | 12 |

# Results

## Data Screening

The data screening process was the same as Study 1.

## Physiological Startle Measures

### Heart Rate

For expected and unexpected events, the mean heart rate scores were calculated using LabChart7. For engine failure events, the heart rate scores were calculated for one minute after the event started, for each pilot, separately. One minute was chosen as following the engine failure, pilots took a between one to four minutes to land. For the aerodynamic stall events, the mean heart rate for the 30 seconds after the beginning of the event were taken for each pilot, separately. Thirty seconds were used as the two stall events took between 10 and 60 seconds. For the baseline heart rate, the mean heart rate of the five central (1 minute after start of file until the sixth minute) minutes of the baseline flight, were calculated for each pilot separately.

#### *Engine Failure*

*Mean Heart Rate*

The unexpected engine failure elicited the highest mean heart rate ($M$ = 98.25bpm, $SD$ =14.74), compared to the expected engine failure ($M$ = 93.74bpm, $SD$=15.88), and the baseline flight ($M$ = 89.24bpm, $SD$ = 12.90) (Figure 12). A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in mean heart rate over the different engine failure flight types (expected and unexpected), and the baseline heart rate flight in pilots. Boxplot diagrams identified one outlier in both engine failure flights. These results were excluded as they appear to be due to a measurement malfunction where the pilots heart rate in all conditions was over 150 beats per minute. After removal two further outliers appeared which were; a low heart rate (approximately 60bpm), and a high heart rate (approximately 120bpm). These heart rates can be considered within the normal physiological range and were therefore kept for the analysis (Froelicher & Myers, 2007). The data were normally distributed for each flight, as assessed by the Shapiro-Wilk test ($p$ >.05 ). The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $\chi^2(2) = 0.898$, $p = .359$. There were statistically significant differences in mean heart rates over the expected and unexpected engine failures and the baseline flight, $F(2,40) = 15.902$, $p < .001$, partial $\eta^2 = .443$. Post hoc analysis with a Bonferroni adjustment revealed that heart rate was statistically significantly higher for the unexpected compared to both the expected engine failure (4.51bpm, 95% CI [0.074, 8.935], $p$ =.045), and the baseline flight (9.01bpm, 95% CI [4.455, 13.563], $p < .0005$). The heart rate was also significantly higher for the expected engine failure compared to baseline flight (4.50bpm, 95% CI [1.057, 7.952], $p < .0005$). The ANOVA was significant with the outlier included and continued to be significant after the removal.

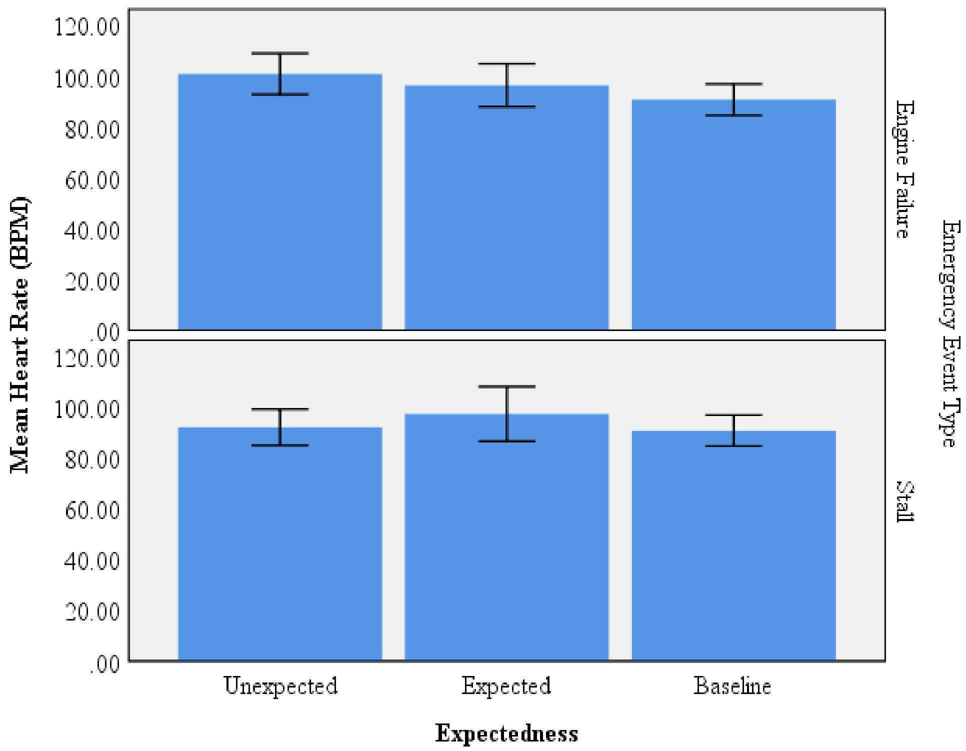



*Figure 12.* The mean heart rates (BPM) for pilot participants, for the expected and unexpected engine failure and stall flights, as well as the baseline flight, with standard error bars (+/-2 SE).

*Difference from baseline*

The mean increase in heart rate from baseline in the expected engine failure was smaller (*M* = 4.50bpm, *SD* = 6.05) compared to the unexpected engine failure (*M* = 9.01bpm, *SD* = 7.99). A difference score was obtained by subtracting the unexpected engine failure difference from the expected engine failure difference. The difference score was normally distributed as shown by a Shapiro-Wilk test ($p < .05$), there was no outliers as assessed by a boxplot. A paired t-test indicated that this was a statistically significant mean difference of 4.50bpm, 95% CI[0.97, 8.04], $t(21) = 2.656$, $p = .015$, $d = 0.58$.

***Aerodynamic Stall***

*Mean heart rate*

Pilots had a higher heart rate in the expected stall flight (*M* = 97.25bpm, *SD* = 22.14) compared to the unexpected stall flight (*M* = 92.02bpm, *SD* =14.77) and the baseline flight (*M* = 90.70bpm, *SD* = 14.33) (Figure 12). A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in mean heart rate between the two different stall flight types (expected and unexpected), and the baseline heart rate flight. There were no statistically significant differences in mean heart rate between the different stall flights and the baseline flight.

*Difference from baseline*

The mean increase in heart rate from baseline in the expected stall was larger (*M* = 4.93bpm, *SD* = 6.96) compared to the unexpected stall (*M* = 1.27bpm, *SD* = 11.98). The difference score was obtained by subtracting the unexpected stall heart rate increase from the expected stall heart rate increase. The difference score was not normally distributed as shown by a Shapiro-Wilk test ($p > .05$), there was two outliers (one extreme and one normal) as assessed by a boxplot. Removal of both outliers led to a normal distribution and no outliers as assessed by boxplots. A paired t-test indicated that this was a statistically significant mean difference of 4.78bpm, 95% CI[0.63, 8.94], $t(13) = 2.489$, $p = .027$, $d = 0.67$. The paired t-test was significant with both the outliers included in the analysis, as well as only the extreme outlier excluded.

**Pupil Dilation**

The Gazepoint software tracks pupil size in pixels (pupil dilation) for both eyes separately over the course of the recording. Using the eye-tracking video it was determined what time the event started and ended. These times were used to calculate separate pupil dilation averages for the left pupil and the right pupil, before and during each event, for each pilot. The average of the left and right pupil was calculated for each participant, separately, producing before event and during event pupil dilation averages for each flight. The difference between these were calculated by calculating the average pupil dilation during the event minus average pupil dilation before the event. The difference was compared between expected and unexpected flights. Comparing the differences controlled for the fact that to keep the flights variable they were programmed at different times of the day which meant different amounts of sunlight. For example the Glentanner (baseline) flight was programmed early in the morning therefore the lighting could have affected the pupil dilation.

***Engine Failure***

*Expected:* As shown in Figure 13, participants on average had a smaller pupil dilation before the expected engine failure ($M = 17.18$ pixels, $SD = 3.02$) and larger pupil dilation during the expected engine failure ($M = 18.55$ pixels, $SD = 3.15$). The difference (during engine failure average pupil dilation – before engine failure pupil dilation) pupil dilation data were assessed and it was found that there was one outlier but the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. The outlier was examined and it was determined to be plausible. A statistically significant mean change in pupil dilation of 1.38 pixels was found, 95% CI[1.014, 1.738], $t(16) = 8.059$, $p < .0005$, $d$

= 1.95. A paired t-test was completed with the outlier removed and with the outlier in place both resulted in significant tests, therefore the outlier was included

*Unexpected:* As shown in Figure 13, participants on average had a smaller pupil dilation before the unexpected engine failure ($M$ = 16.17 pixels, $SD$ = 2.84) and larger pupil dilation during the unexpected engine failure ($M$ = 18.11 pixels, $SD$ = 3.25). The difference pupil dilation data were assessed and it was found that there was one outlier but the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. The outlier was examined and it was determined to be plausible. A paired t-test completed with the outlier removed and with the outlier in place both resulted in significant tests, therefore the outlier was included. A statistically significant mean change in pupil dilation of 1.94 pixels was found, 95% CI[1.485, 2.392], $t(16) = 9.062$, $p < .0005$, $d = 2.20$.

*Change in pupil dilation:* As shown in Figure 13, during the unexpected engine failure pilots had a larger increase in pupil dilation ($M$ = 1.83 pixels, $SD$ = 0.72), compared to the expected engine failure ($M$ = 1.32 pixels, $SD$ = 0.50). The change in pupil dilation in the expected engine flight was subtracted from the change in pupil dilation in the unexpected engine failure flight. The resulting difference variable was analysed for normality and outliers. The difference pupil dilation data were assessed and it was found that there were two outliers, one extreme outlier, and one standard outlier, the data were not normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p = .050$), respectively. Removal of both the outliers, resulted in a normal distribution of data with no outliers, as assessed by boxplot and Shapiro-Wilk test ($p >.05$), respectively. A paired t-test indicated there was a statistically significant larger mean increase in pupil dilation of 0.54 pixels in the unexpected engine failure, 95% CI[0.177, 0.837], $t(14) = 3.296$, $p = .005$, $d = 0.85$. It is important to note that before the removal of the outliers the paired t-test was only marginally statistically significant (*p* = .057). Only removing the extreme outlier also resulted in a non-significant paired t-test (*p* = .091).

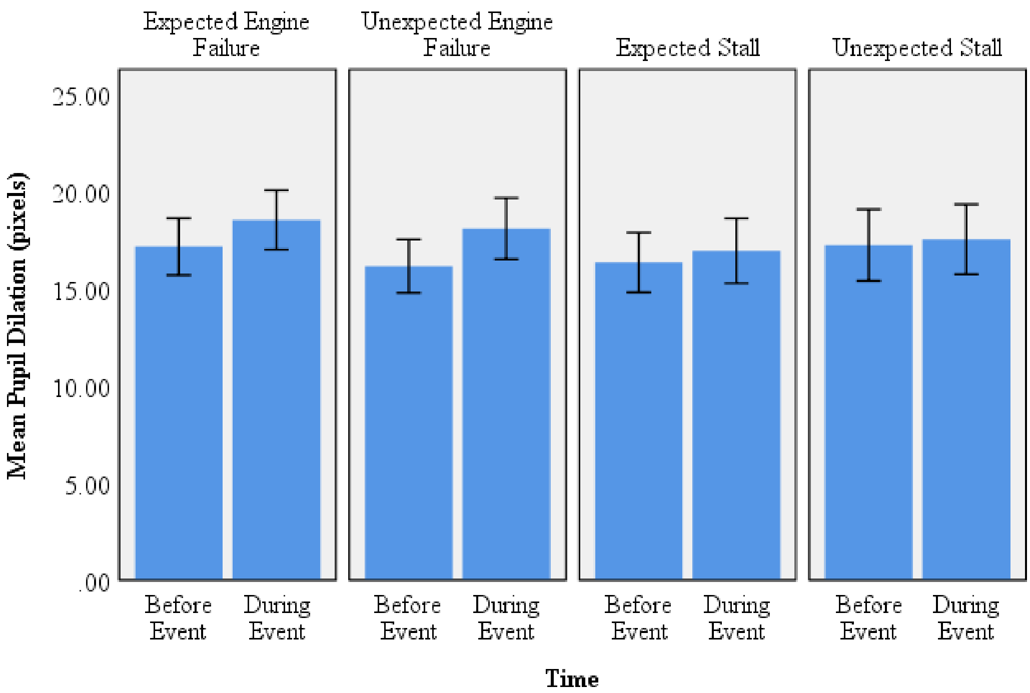



*Figure 13.* Mean pilot pupil pixels (dilation) for before and during each expected and unexpected stall and engine failure with standard error bars (+/-2 SE).

**Aerodynamic Stall**

*Expected:* Pilots on average had a smaller pupil dilation before the practice stall (*M* = 16.36 pixels, *SD* = 2.79) and slightly larger pupil dilation during the practice stall (*M* = 16.96 pixels, *SD* = 2.26) (Figure 13). There was no significant difference between pupil dilation before and after the practice stall.

*Unexpected:* As shown in Figure 13, participants on average had a smaller pupil dilation before the unexpected stall (*M* = 17.10 pixels, *SD* = 3.42) and larger pupil dilation during the unexpected stall (*M* = 17.75, *SD* = 3.30). The difference pupil dilation data were assessed and it was found that there were two outliers, one extreme and another standard, and the data were not normally distributed, as assessed by boxplot and Shapiro-Wilk test (*p* = .001), respectively. The extreme outlier was examined and it was determined to not be plausible when comparing to all other pupil dilation data. It was excluded from the analysis, the other outlier was kept as it was considered plausible when compared to other difference scores. After excluding the extreme outlier, the data were assessed, and found to be normally distributed (Shapiro-Wilk test ($p > 0.05$)). However, the secondary analyses of the boxplot showed two further outliers which were assessed and were found to be plausible. Paired t-test completed on data with the two further outliers removed and with the outliers in place both resulted in significant results, so it was decided to report the results with the included outliers. A paired t-test indicated aa statistically significant mean difference of 0.65 pixels, 95% CI[0.248, 1.051], $t(11) = 3.562$, $p = .004$, $d = 1.03$.

*Change in pupil dilation: T*he unexpected stall had on average a slightly larger change in pupil dilation ($M = 0.65$ pixels $SD = 0.63$) compared to the expected practice stall ($M = 0.61$ pixels, $SD = 1.11$). A paired t-test indicated that this difference was not significant.

**Information Gathering and Processing**

Eye tracking data were used to calculated duration of fixation on AOI. AOI were the different flight instruments displayed on the PFD and the MFD e.g. the altimeter (Appendix I). The percentage time spent looking at each of the AOI was recorded from the beginning to the end of each event (Appendix I). The percentage of time pilots time spent looking inside the cockpit (head down display) at the flight display (PFD and MFD) was calculated.

As Appendix I shows the instruments on the PFD and MFD were split into 10 AOI; Airspeed indicator; Attitude Indicator; Altimeter; Engine Instruments; Turn Coordinator, Directional Gyro, Vertical Speed Indicator, GPS, Navigational Instruments (grouping of VOR1, VOR2, and ADF), and the entire flight display (MFD and PFD). The differences between the percentage time viewed in the unexpected compared to the expected engine failure were calculated for each AOI (expected – unexpected). Before inferential testing, this difference data was assessed for outliers and the normality.

Due to multiple testing occurring, an adjustment was made in so that only analyses where the p value as under 0.025 were considered significant. This value was chosen to prevent Type I and Type II errors.

***Engine Failure***

*Airspeed Indicator*

As shown in Figure 14, pilots spent a larger percentage of their time observing the airspeed indicator in the expected engine failure practice ($M$ = 1.85%, $SD$ = 1.36) as opposed to the unexpected engine failure emergency ($M$ = 0.79%, $SD$ = 0.65). The difference data were assessed and it was found that there were no outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p$ >.05), respectively. A paired t-test indicated that there was a statistically significant mean difference of 1.06%, 95% CI[0.380, 1.737], $t$(16) = 3.308, $p$ = .004, $d$ = 0.80.

*Attitude Indicator*

As shown in Figure 14, pilots spent a larger percentage of their time observing the attitude indicator in the expected engine failure practice ($M$ = 0.28%, $SD$ = 0.20) as opposed to the unexpected engine failure emergency ($M$ = 0.15%, $SD$ = 0.11).. A paired t-test indicated that there was no statistically significant differences ($p$ = .028).

*Altimeter*

As shown in Figure 14, the untransformed data showed that pilots spent a larger percentage of their time looking at the altitude meter in the expected engine failure flight ($M$ = 0.84%, $SD$ = 0.91) compared to in the unexpected engine failure ($M$ = 0.22%, $SD$ = 0.18). The difference data were assessed and it was found that there were 4 outliers and the data were not normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p$ = .001), respectively. Histograms show that data were moderately positively skewed. Therefore, a square root transformation was conducted. The transformed difference data were assessed and it was found that there were no outliers and the data were normally distributed for each flight, as assessed by boxplot and Shapiro-Wilk test ($p$ >.05), respectively. A paired t-test performed on the transformed data showed that this was a significant difference, $t$(16) =3.304, $p$ = .004. The paired t-test completed on the untransformed data was also significant.

*GPS*

As shown in Figure 14, the untransformed data showed that pilots spent a greater percentage of their time looking at the GPS in the expected engine failure flight ($M$ = 0.53%, $SD$ = 0.61) compared to in the unexpected engine failure ($M$ = 0.29%, $SD$ = 0.18). The difference data were assessed and it was found that there were two outliers and the data were not normally distributed, as assessed by boxplot and Shapiro-Wilk test (p < .0005), respectively. Histograms show that the both the expected and unexpected data were moderately positively skewed. Therefore, a square root transformation was conducted on both the percentage time expected and unexpected attitude indicator AOI data. The difference of these two scores was computed. The square root difference data were assessed and it was found that there were two outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p$ >.05), respectively. The outliers were retained as they very close to the next highest acceptable value (0.57 and 0.56 versus 0.53). A paired t-test performed on the transformed data showed that this was a significant difference, $t$(16) =3.104, $p$ = .007. The paired t-test completed on the untransformed data were also significant.

*Turn Coordinator*

As shown in Figure 14, the untransformed data showed that participants spent a larger percentage of their time looking at the turn coordinator in the expected engine failure flight ($M$ = 0.24%, $SD$ = 0.25) compared to in the unexpected engine failure ($M$ = 0.12%, $SD$ = 0.10). The difference data were assessed and it was found that there were two outliers (Pilot 20; -0.51, and pilot 10, -0.57) and the data were not normally distributed for each flight, as assessed by boxplot and Shapiro-Wilk test ($p$ = .014), respectively. These outliers were considered as plausible values, therefore, a square root transformation was conducted on both the percentage time expected and unexpected turn coordinator AOI data. The difference of these two scores was computed. The square root difference data were assessed and it was found that there were no outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p$ >.05), respectively. A paired t-test performed on the transformed data showed that this was a significant difference, $t$(16) =2.944, $p$ = .010. The paired t-tests completed on the untransformed data with and without the outliers were also significant.

*Directional Gyro*

As shown in Figure 14, the untransformed data showed that pilots spent a longer percentage of their time looking at the directional gyro in the expected engine failure flight ($M$ = 0.14%, $SD$ = 0.14) compared to in the unexpected engine failure ($M$ = 0.07%, $SD$ = 0.06). The difference data were assessed and it was found that there was one outlier and the data were not normally distributed for each flight, as assessed by boxplot and Shapiro-Wilk test ($p$ = .036), respectively. The data were positively skewed, therefore, a square root transformation was conducted on both the percentage time expected and unexpected directional gyro AOI data. The difference of these two scores was computed. The square root difference data were assessed and it was found that there were no outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p$ >.05), respectively. A paired t-test performed on the transformed data showed that this was a significant difference, $t$(16) =2.862, $p$ = .011. Paired t-tests completed on the untransformed data with and without the outlier were also significant.

*Vertical Speed Gauge*

Pilots spent a larger percentage of their time observing the vertical speed indicator in the expected engine failure practice (*M* = 0.21%, *SD* = 0.22) as opposed to the unexpected engine failure (*M* = 0.10%, *SD* = 0.07) (Figure 14). A paired t-test showed that this difference was not significant.

*Navigational Instruments (grouping of VOR1, VOR2, and ADF)*

Pilots spent a greater percentage of their time observing the navigational instruments in the expected engine failure practice (*M* = 0.25%, *SD* = 0.22) as opposed to the unexpected engine failure (*M* = 0.15%, *SD* = 0.11) (Figure 14). A paired t-test showed that this difference was not significant.

*Engine Instruments*

Pilots spent a larger percentage of their time observing the engine instruments in the expected engine failure practice (*M* = 0.12%, *SD* = 0.18) as opposed to the unexpected engine failure (*M* = 0.07%, *SD* = 0.10) (Figure 14). A paired t-test showed that this difference was not significant.

*Inside*

As shown in Figure 15, pilots spent a larger percentage of their time observing the PFD and MFD displayed on the LCD screen inside the cockpit in the expected engine failure practice (*M* = 6.51%, *SD* = 2.17) as opposed to the unexpected engine failure emergency (*M* = 3.73%, *SD* = 1.44). The difference data were assessed and it was found that there were no outliers and the data were normally distributed for each flight, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. There was a statistically significant mean difference of 2.82%, 95% CI[1.333, 4.228], $t(16) = 4.073$, $p = .001$, $d = 0.99$.

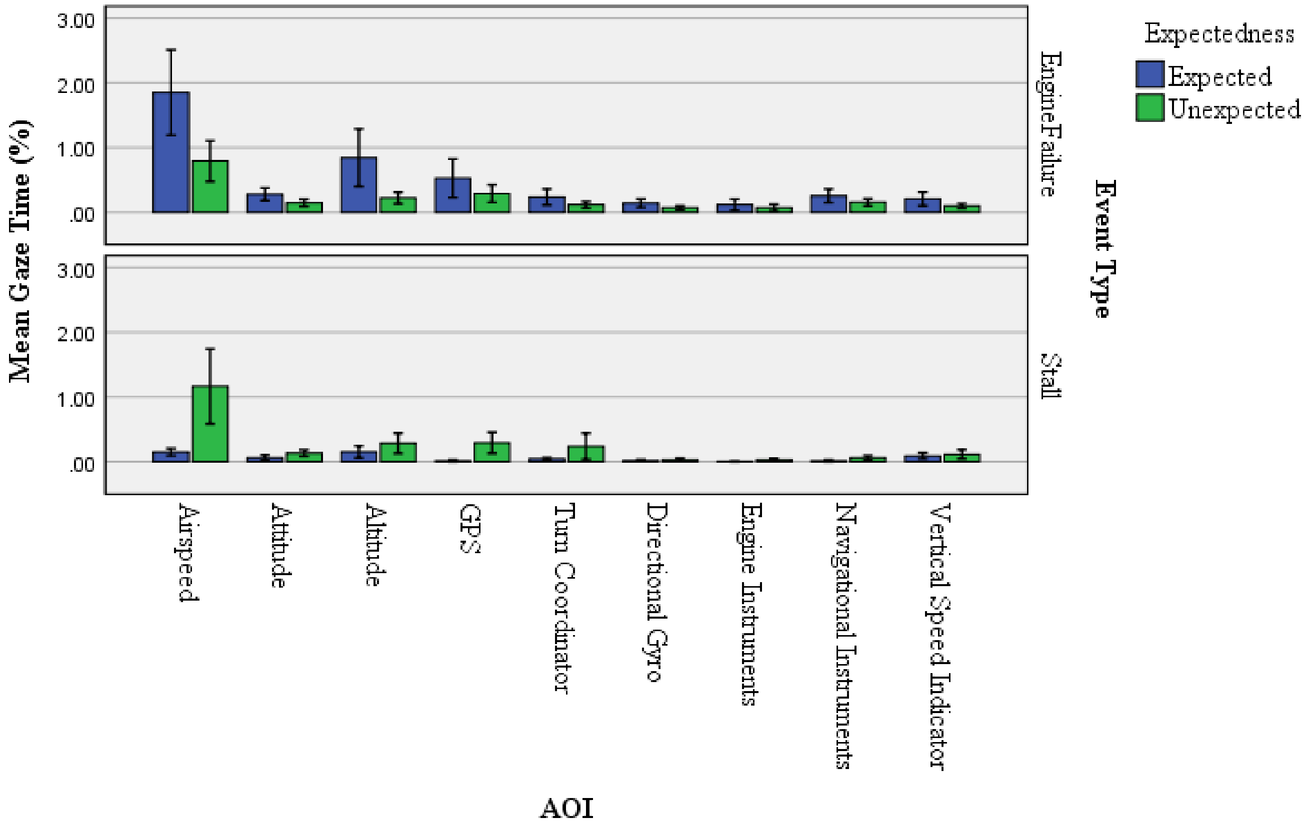



*Figure 14.* Mean percentage time pilots spent looking at the different AOI (flight instruments) from the PFD and MFD during the expected and unexpected engine failures and aerodynamic stall flights with standard error bars (+/-2 SE).

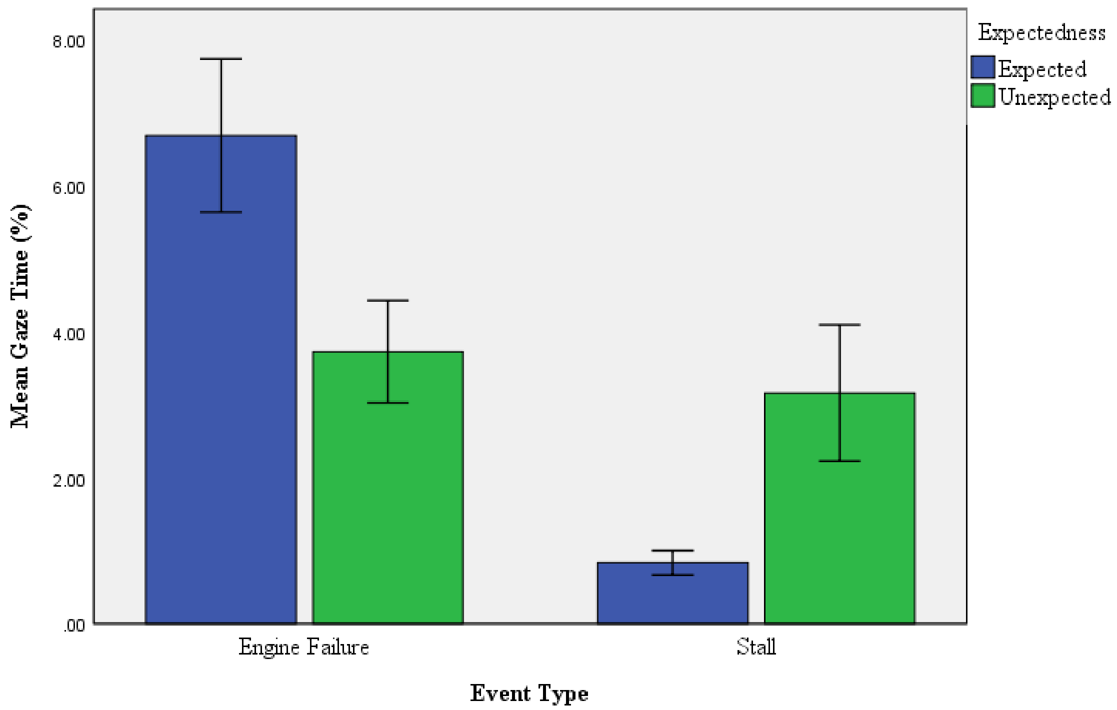



*Figure 15.* Mean percentage time pilots spent looking at the combined PFD and MFD during the expected and unexpected engine failure and aerodynamic stall flights with standard error bars (+/-2 SE).

***Stall***

*Airspeed Indicator*

As shown in Figure 14, pilots spent a larger percentage of their time observing the airspeed indicator during the unexpected stall (*M* = 1.16%, *SD* = 1.04) as opposed to the expected stall (*M* = 0.15%, *SD* = 0.10) The difference data were assessed and it was found that there were no outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p$ >.05), respectively. There was a statistically significant mean difference of 1.01%, 95% CI[0.399, 1.629], $t(12) = 3.594$, $p = .004$, $d = 1.00$.

*Attitude Indicator*

As shown in Figure 14, pilots spent a larger percentage of their time observing the attitude indicator during the unexpected stall (*M* = 0.13%, *SD* = 0.09) as opposed to the expected stall (*M* = 0.06%, *SD* = 0.07). The difference data were assessed and it was found that there were no outliers and the data were normally distributed for each flight, as assessed by boxplot and Shapiro-Wilk test ($p$ >.05), respectively. There was a statistically significant mean difference of 0.07%, 95% CI[0.017, 0.125], $t(12) = 2.887$, $p = .014$, $d = 0.80$.

*Altimeter*

Pilots spent a smaller percentage of their time observing the altimeter speed indicator during the expected stall (*M* = 0.15%, *SD* = 0.16) as opposed to the unexpected stall (*M* = 0.28%, *SD* = 0.28) (Figure 14). A paired t-test showed that this difference was not significant.

*GPS*

As shown in Figure 14, pilots spent a larger percentage of their time looking at the GPS in the unexpected stall (*M* = 0.29%, *SD* = 0.29) compared to in the expected stall (*M* = 0.01%, *SD* = 0.02). The difference data were assessed and it was found that there was one outlier and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test

($p > .05$), respectively. The outlier was assessed and decided to be plausible, and thus was kept in the analysis. A paired t-test performed showed that there was a significant difference of 0.28%, 95% CI[0.107, 0.450], $t(12) = 3.539$, $p = .004$, $d = 0.98$.

*Turn Coordinator*

As shown in Figure 14, the untransformed data showed that participants spent a larger percentage of their time looking at the turn coordinator in the unexpected stall flight ($M$ = 0.24%, $SD$ = 0.36) compared to during the expected stall ($M$ = 0.04%, $SD$ = 0.04). The difference data were assessed and it was found that there were two extreme outliers and the data were not normally distributed (positively skewed, as assessed by boxplot and Shapiro-Wilk test ($p < .0005$), respectively. These outliers were considered as plausible values, therefore, a square root transformation was conducted on both the percentage time expected and unexpected turn coordinator AOI data. The difference of these two scores was computed. The square root difference data were assessed and it was found that there were no outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p >.05$), respectively. A paired t-test performed on the transformed data showed that there was a significant difference, $t(12) =2.555$, $p = .025$. The paired t-test completed on the untransformed data excluding the outliers was also significant. The paired t-test completed on the untransformed data with the outliers was not significant.

*Engine Instruments*

As shown in Figure 14, the untransformed data indicates that participants spent a slightly greater percentage of their time looking at the engine instruments in the unexpected stall ($M$ = 0.03%, $SD$ = 0.04) compared to during the expected stall ($M$ = 0.002%, $SD$ = 0.004). The difference data were assessed and it was found that there were no outliers and the data were not normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p = .007$), respectively. Observation of a frequency histogram revealed a positively skewed distribution therefore a square root transformation was applied to both the percentage time expected and unexpected engine instrument group AOI data. The difference of these two scores was computed. The square root difference data were assessed and it was found that there were no outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p$ >.05), respectively. A paired t-test performed on the transformed data indicated that there was a significant difference, $t(12)$ =2.770, $p$ = .017. The paired t-test completed on the untransformed data was also significant.

*Navigational Instruments (grouping of VOR1, VOR2, and ADF)*

As shown in Figure 14, pilots spent a larger percentage of their time looking at the navigational instrument group in the unexpected stall flight ($M$ = 0.06%, $SD$ = 0.06) compared to in the expected stall flight ($M$ = 0.01%, $SD$ = 0.03). The difference data were assessed and it was found that there was no outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p$ > .05), respectively. A paired t-test performed showed that there was a significant difference of 0.05%, 95% CI[0.012, 0.022], $t(12)$ = 4.069, $p$ = .002, $d$ = 1.13.

*Directional Gyro*

Participants spent a similar percentage of their time observing the directional gyro in the expected stall practice ($M$ = 0.02%, $SD$ = 0.03) and the unexpected stall ($M$ = 0.03%, $SD$ = 0.04) (Figure 14). A paired t-test showed that this difference was not significant.

*Vertical Speed Gauge*

Participants spent a slightly lower percentage of their time observing the vertical speed gauge in the expected stall practice ($M$ = 0.09%, $SD$ = 0.08) as opposed to the unexpected stall ($M$ = 0.12%, $SD$ = 0.13) (Figure 14). A paired t-test indicated that this difference was not significant.

*Inside*

As shown in Figure 14, pilots spent a larger percentage of their time observing the PFD and MFD displayed on the LCD screen inside the cockpit during the unexpected stall event ($M$ = 3.16%, $SD$ = 1.68) as opposed to the expected stall e ($M$ = 0.83%, $SD$ = 0.30). The difference data were assessed and it was found that there were no outliers and the data were normally distributed, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. A paired t-test indicated that there was a statistically significant mean difference of 2.33%, 95% CI[1.320, 3.332], $t(12) = 5.037$, $p < .0005$, $d = 1.40$.

## Performance

***Engine Failure***

*Crash*

In the unexpected engine failure 54.5% (12 of 22) of the pilots landed safely and 45.5% (10 of 22) of the pilots either crashed or attempted to land on the water. No pilots crashed in the expected engine failure crash.

*Time to pull back on throttle to prevent engine surges affecting glide*

The time to pull back on the throttle was calculated from when the engine failure occurred to when the throttle was completely closed. This is done during an engine failure to prevent any engine surges affecting the planes glide (CAA, 2012). In both flights six pilots out of 19 (31.60%) failed to retard the throttle after engine failure. These pilots were excluded from the analysis. In the unexpected engine failure flight pilots took longer to retard the throttle ($N$ = 12, $M$ = 3.42 seconds, $SD$ = 2.11), than in the expected engine failure condition ($N$ = 12, $M$ = 2.67 seconds, $SD$ = 2.42). A paired t-test showed that this difference was not significant.

***Stall***

*Stall Recovery Time*

The time to recover from the expected practice stall was calculated from the time that the pilot retarded the throttle, and the airspeed was reduced to 60 knots, then after the stall occurred till the plane returned to level flight. The time to recover from the unexpected stall was calculated from the time that the plane reached 500ft till the time that the pilot successfully passed the wind shear and reached 550ft. On average pilots took longer to recover in the unexpected stall flight (*N* = 12, *M* = 36.98 seconds, *SD* = 20.49) compared to the expected stall flight (*N = 12*, *M* = 17.95, *SD* = 3.47). Due to differences in calculation method, and flight configuration, recovery time is not comparable between the unexpected and the expected stall flights.

*Altitude lost*

On average pilots lost more altitude in the expected stall flight (*N* = 12, *M* = 215.00 feet, *SD* = 73.31) compared to the unexpected stall flight (*N* = 12, *M* = 59.62, *SD* = 67.03). The altitude lost is not comparable between the two due to the fact that the tail wind change was only between 500 and 550 feet, therefore when the pilots dropped below 500feet in the unexpected condition, the head wind picked up, this minimised the altitude loss. This probably explains why the mean altitude lost was only around 50 feet.

*Applying the correct recovery*

In the unexpected stall flight 30.8% (4 of 13) pilots incorrectly pulled back on the throttle after hearing the stall warning. As well as this in the unexpected stall flight 38% (5 of 13) pilots incorrectly did not lower the nose of the plane and pitch it downwards in response to the stall horn

*Time to apply maximum power*

As in the Casner et al. (2012) study, the time to apply maximum power was examined, however all pilots already had the throttle at maximum power before the stall occurred.

**Performance and Heart Rate**

Pilot 15 was excluded from all heart rate analyses due to having abnormal heart rate measurements.

***Engine Failure***

Pilots that landed safely in the unexpected engine failure flight had a higher heart rate ($N$ = 11, $M$ = 100.57, $SD$ = 14.31), than those that crashed in the unexpected engine failure flight ($N$ = 10, $M$ = 95.69, $SD$ = 15.53). An independent t-test showed that this difference was not significant.

***Stall***

*Pulled back on throttle*

Pilots who incorrectly pulled back on the throttle after the stall horn sounded showed higher heart rates ($N = 4$, $M$ = 91.03, $SD$ = 7.51), than pilots who did not move the throttle ($N = 9$, $M$ = 89.25, $SD$ = 12.84). An independent t-test showed that this difference was not significant.

*Pitched the nose downwards*

Pilots who correctly pitched the nose downwards after the stall horn sounded had lower heart rates ($N = 9$, $M$ = 87.29, $SD$ = 4.11), than pilots who did not pitch the nose downwards ($N$ = 4, $M$ = 94.21, $SD$ = 8.75). An independent t-test showed that this difference was not significant.

**Performance and Pupil Dilation.**

***Engine Failure***

*Crash landing or safely landing in unexpected engine failure and heart rate*

Participants that landed safely in the unexpected engine failure flight had slightly smaller pupil dilation ($N = 12$, $M = 18.04$, $SD = 2.95$), than those that crashed in the unexpected engine failure flight ($N = 10$, $M = 18.20$, $SD = 3.77$). An independent t-test showed that this difference was not significant.

Participants that landed safely in the unexpected engine failure had a slightly larger increase in pupil dilation from before the engine failure to during the engine failure ($N = 12$, $M = 1.97$, $SD = 0.96$), compared to those that crashed ($N = 10$, $M = 1.91$, $SD = 0.86$). An independent t-test showed that this difference was not significant.

Stall

In the stall flights, one pilot had extreme outlier data for pupil dilation indicating a likely measurement error for one of the stall flights; his data was excluded from the analyses.

*Pull back on throttle*

Pilots who incorrectly pulled back on the throttle had a larger pupil dilation during the unexpected stall ($N = 4$, $M = 18.33$, $SD = 4.50$), compared to pilots who did not pull back on the throttle ($N = 8$, $M = 17.47$, $SD = 2.84$). An independent t-test showed that this difference was not significant.

Pilots who incorrectly pulled back on the throttle had a smaller increase in pupil dilation from before the stall to during the stall ($N = 4$, $M = 0.48$, $SD = 0.35$), compared to pilots who did not pull back on the throttle ($N = 8$, $M = 0.73$, $SD = 0.74$). An independent t-test showed that this difference was not significant.

*Pitched nose downwards*

Pilots who correctly pitched the nose downwards after the stall horn showed a smaller pupil dilation (*N* = 4, *M* = 17.58, *SD* = 2.75), compared to pilots who did not pitch the nose of the plane down (*N = 8*, *M* = 18.10, *SD* = 4.69). An independent t-test showed that this difference was not significant.

Pilots who correctly pitched the nose downwards after the stall horn sounded showed smaller increase in pupil dilation (*N = 8*, *M* = 0.52, *SD* = 0.51), whereas pilots that did not pitch the nose downwards showed a larger increase in pupil dilation (*N = 4, M* = 0.91, *SD* = 0.86). An independent t-test showed that this difference was not significant.

**Performance and AOIS**

***Engine Failure***

*Airspeed*

Pilots that landed safely after the unexpected engine failure spent a larger percentage of their time looking at the airspeed indicator (*M* = 0.94, *SD* = 0.76), compared to pilots who crashed (*M* = 0.63, *SD* = 0.50) (Figure 16). An independent t-test revealed that this was not a significant difference.

*Attitude Indicator*

Pilots that landed safely after the unexpected engine failure spent a greater percentage of their time looking at the attitude indicator (*M* = 0.17, *SD* = 0.13), compared to pilots who crashed (*M* = 0.13, *SD* = 0.10) (Figure 16). An independent t-test revealed that this was not a significant difference.

*Altimeter*

Pilots that landed safely after the unexpected engine failure spent a greater percentage of their time looking at the altimeter (*M* = 0.26, *SD* = 0.21), compared to pilots who crashed

(*M* = 0.18, *SD* = 0.15) (Figure 16). An independent t-test revealed that this was not a significant difference.

*Navigational Instruments*

Pilots that landed safely after the unexpected engine failure spent a smaller percentage of their time looking at the navigational instruments (*M* = 0.11, *SD* = 0.06), compared to pilots who crashed (*M* = 0.20, *SD* = 0.14) (Figure 16). An independent t-test revealed that this was not a significant difference.

*GPS*

Pilots that landed safely after the unexpected engine failure spent a smaller percentage of their time looking at the GPS (*M* = 0.27, *SD* = 0.17), compared to pilots who crashed (*M* = 0.31, *SD* = 0.39) (Figure 16). An independent t-test revealed that this was not a significant difference.

*Engine Instruments*

Pilots that landed safely after the unexpected engine failure spent a similar percentage of their time looking at the engine instruments (*M* = 0.08, *SD* = 0.13), as the pilots who crashed (*M* = 0.06, *SD* = 0.08) (Figure 16). An independent t-test revealed that there were no significant differences.

*Turn Coordinator*

Pilots that landed safely after the unexpected engine failure spent a similar percentage of their time looking at the turn coordinator (*M* = 0.11, *SD* = 0.06), as the pilots who crashed (*M* = 0.12, *SD* = 0.14) (Figure 16). An independent t-test revealed that there were no significant differences.

*Directional Gyro*

Pilots that landed safely after the unexpected engine failure spent a similar percentage of their time looking at the directional gyro (*M* = 0.08, *SD* = 0.06), as the pilots who crashed

(*M* = 0.06, *SD* = 0.06) (Figure 16). An independent t-test revealed that this was not a significant difference.

*Vertical Speed Indicator*

Pilots that landed safely after the unexpected engine failure spent a similar percentage of their time looking at the vertical speed indicator (*M* = 0.09, *SD* = 0.08), as the pilots who crashed (*M* = 0.10, *SD* = 0.08) (Figure 16). An independent t-test revealed that there were no significant differences.

*Inside Cockpit (Head down display)*

Pilots that landed safely after the unexpected engine failure spent a smaller percentage of their time looking inside the cockpit at the screen displaying the MFD and the PFD (*M* = 3.58, *SD* = 1.02), compared to pilots who crashed (*M* = 3.89, *SD* = 1.87) (Figure 17). An independent t-test revealed that this was not a significant difference.

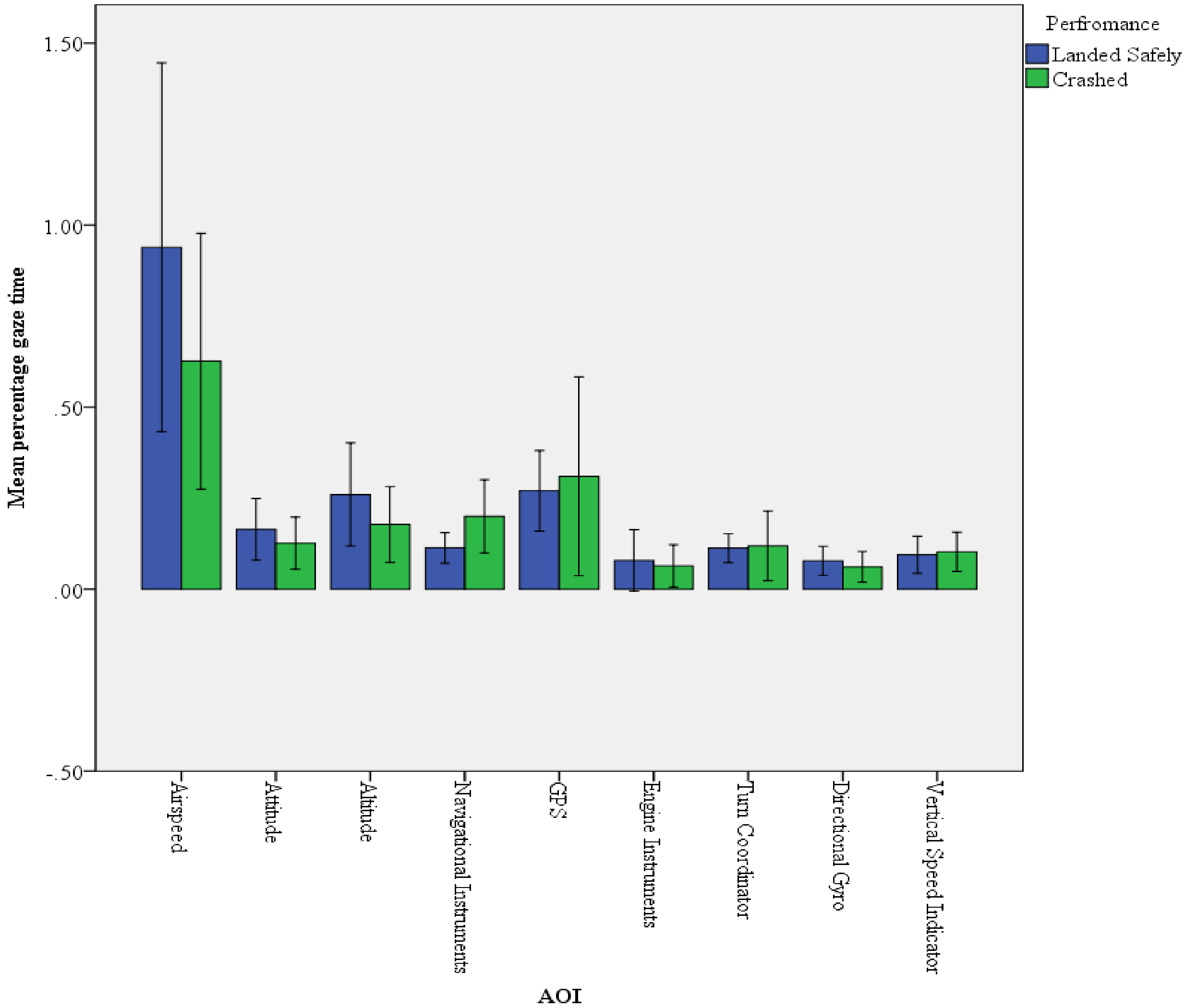



*Figure 16.* Mean percentage gaze time for each different AOI (flight instrument) for pilots who crashed or landed safely in the unexpected engine failure flight with standard error bars (+/-2SE).

***Stall***

*Airspeed*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a larger percentage of their time looking at the airspeed indicator ($M$ = 1.61, $SD$ = 0.83), compared to pilots who responded incorrectly ($M$ = 1.52, $SD$ = 1.13). An independent t-test revealed that this was not a significant difference.

Pilots that correctly left the throttle at full power following the stall spent a larger percentage of their time looking at the airspeed indicator ($M$ = 1.52, $SD$ = 0.83), compared to pilots who responded incorrectly by pulling back the throttle (*M* = 1.16, *SD* = 1.13). An independent t-test revealed that this was not a significant difference.

*Attitude Indicator*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a slightly larger percentage of their time looking at the attitude indicator (*M* = 0.16, *SD* = 0.08), compared to pilots who responded incorrectly (*M* = 0.12, *SD* = 0.10). An independent t-test revealed that this was not a significant difference.

Pilots that correctly left the throttle at full power following the stall spent a smaller percentage of their time looking at the attitude indicator (*M* = 0.14, *SD* = 0.96), compared to pilots who responded incorrectly by pulling back the throttle (*M* = 0.16, *SD* = 0.06). An independent t-test revealed that this was not a significant difference.

*Altimeter*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a larger percentage of their time looking at the altimeter (*M* = 0.46, *SD* = 0.25), compared to pilots who responded incorrectly (*M* = 0.10, *SD* = 0.18). The data from the correct pilot group had no outliers and their data was normally distributed. The data from the incorrect pilots was not normally distributed and they had 1 extreme outlier and 1 standard outlier. This is likely due to the very small sample numbers. It was decided not to transform a distribution based on very small numbers. An independent samples t-test showed statistically significant mean difference of 0.36%, 95% CI[0.070, 0.655], $t(10) = 2.763$, $p = .020$, $d = 0.771$.

Pilots that correctly left the throttle at full power following the stall spent a smaller percentage of their time looking at the altimeter (*M* = 0.28, *SD* = 0.28), compared to pilots who responded incorrectly by pulling back the throttle (*M* = 0.38, *SD* = 0.34). An independent t-test revealed that this was not a significant difference.

*Navigational Instruments*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a larger percentage of their time looking at the navigational instruments ($M$ = 0.07, $SD$ = 0.07), compared to pilots who responded incorrectly ($M$ = 0.04, $SD$ = 0.04). An independent t-test revealed that this was not a significant difference.

Pilots that correctly left the throttle at full power following the stall spent a larger percentage of their time looking at the navigational instruments ($M$ = 0.07, $SD$ = 0.06), compared to pilots who responded incorrectly by pulling back the throttle ($M$ = 0.03, $SD$ = 0.04). An independent t-test revealed that this was not a significant difference.

*GPS*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a smaller percentage of their time looking at the GPS ($M$ = 0.23, $SD$ = 0.23), compared to pilots who responded incorrectly ($M$ = 0.41, $SD$ = 0.39). An independent t-test revealed that this was not a significant difference.

Pilots that correctly left the throttle at full power following the stall spent a smaller percentage of their time looking at the GPS ($M$ = 0.29, $SD$ = 0.22), compared to pilots who responded incorrectly by pulling back the throttle ($M$ = 0.34, $SD$ = 0.56). An independent t-test revealed that this was not a significant difference.

*Engine Instruments*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a similar percentage of their time looking at the engine instruments ($M$ = 0.03, $SD$ = 0.04), compared to pilots who responded incorrectly ($M$ = 0.03, $SD$ = 0.04).

Pilots that correctly left the throttle at full power following the stall spent a similar percentage of their time looking at the engine instruments ($M$ = 0.03, $SD$ = 0.04), compared to pilots who responded incorrectly by pulling back the throttle ($M$ = 0.03, $SD$ = 0.04).

*Turn Coordinator*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a larger percentage of their time looking at the turn coordinator (*M* = 0.36, *SD* = 0.47), compared to pilots who responded incorrectly (*M* = 0.07, *SD* = 0.07). An independent t-test revealed that this was not a significant difference.

Pilots that correctly left the throttle at full power following the stall spent a larger percentage of their time looking at the turn coordinator (*M* = 0.30, *SD* = 0.43), compared to pilots who responded incorrectly by pulling back the throttle (*M* = 0.08, *SD* = 0.10). An independent t-test revealed that this was not a significant difference.

*Directional Gyro*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a similar percentage of their time looking at the directional gyro (*M* = 0.04, *SD* = 0.05), compared to pilots who responded incorrectly (*M* = 0.03, *SD* = 0.04). An independent t-test revealed that there was no significant differences.

Pilots that correctly left the throttle at full power following the stall spent a similar percentage of their time looking at the directional gyro (*M* = 0.03, *SD* = 0.05), compared to pilots who responded incorrectly by pulling back the throttle (*M* = 0.04, *SD* = 0.05). An independent t-test showed that this was not a significant differences.

*Vertical Speed Indicator*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a larger percentage of their time looking at the vertical speed indicator (*M* = 0.20, *SD* = 0.11), compared to pilots who responded incorrectly (*M* = 0.02, *SD* = 0.04). The data from the correct pilot group had two extreme outliers and their data was normally distributed. The data from the incorrect pilots was not normally distributed and they had 1 extreme outlier. This is likely due to the very small sample numbers. It was decided not to transform a distribution based on very small numbers. An independent samples t-test showed statistically significant mean difference of 0.18%, 95% CI[0.060, 0.300], $t(10) = 3.350$, $p = .007$, $d = 0.616$.

Pilots that correctly left the throttle at full power following the stall spent a larger percentage of their time looking at the vertical speed indicator ($M = 0.14$, $SD = 0.14$), compared to pilots who responded incorrectly by pulling back the throttle ($M = 0.09$, $SD = 0.09$). An independent t-test showed that this was not a significant difference.

*Inside Cockpit (Head down display)*

Pilots that correctly pointed the nose of the plane downwards following the stall horn spent a smaller percentage of their time looking inside the cockpit at the screen displaying the MFD and the PFD ($M = 4.07$, $SD = 1.29$), compared to pilots who responded incorrectly ($M = 2.36$, $SD = 1.54$). An independent t-test revealed that this was not a significant difference.

Pilots that correctly left the throttle at full power following the stall spent a smaller percentage of their time looking inside the cockpit at the screen displaying the MFD and the PFD ($M = 3.24$, $SD = 1.72$), compared to pilots who responded incorrectly by pulling back the throttle ($M = 3.71$, $SD = 1.36$). An independent t-test showed that this was not a significant difference.

# Discussion

## Physiological Measurements

It was hypothesized that pilots would have a higher heart rate and larger pupil dilation during the unexpected flight events (engine failure and stall) compared to the expected flight events and baseline (heart rate) or before the event (pupil dilation). It was also hypothesized that pilots would have a higher heart rate and larger pupil dilation during the expected events compared to the baseline flight or before the events, respectively. Consistent with the hypotheses pilots had significantly higher heart rates in the unexpected engine failure, compared to the expected engine failure, and the baseline flight. The pilots' heart rates in the expected engine failure were also significantly higher than during the baseline flight. In support of the hypothesis, during the unexpected engine failure and the expected engine failure, pilots had significantly larger mean pupil dilations during the event compared to before the event. Furthermore, the unexpected engine failure led to a significantly larger mean increase in pupil dilation than the expected engine failure. These findings indicate that both the expected and unexpected engine failures led to increases in autonomic arousal, and the unexpected engine failure had the largest autonomic response.

When comparing the data with other research, the unexpected engine failure resulted in arousal analogous to startle, whereas the expected engine failure showed arousal consistent with an increased mental workload. The pilots mean heart rate increase compared to the baseline flight in the unexpected engine failure was 9.01bpm. This value is very similar to the mean increase in heart rate found by other research that has investigated physiological responses to an auditory startle stimulus (Chou et al., 2014; Deuter et al., 2012; Holand et al., 1999). Lahtinen, Koskelo, Laitinen, and Leino (2007) found that during extremely difficult tactical manoeuvres in combat missions, pilots' heart rates increased by approximately 10-15bpm. Therefore the data suggest that the mean increase in heart rate found during the unexpected engine failure in the current study was similar to startle and pilots' physiological reactions in high stress situations. Additionally, pupil dilation has been shown to increase during the startle reaction (Bradley et al., 2005; Rivera et al., 2014). During the expected engine failure pilots had a mean increase in heart rate of 4.50bpm. This heart rate increase was similar to the findings of Fallahi et al. (2016) and Grassmann et al. (2017) who investigated the changes in heart rate in response to increasing workload. In support of this, Lahtinen et al. (2007) showed that pilots had an increase of around 5bpm above baseline during the simpler parts of their combat flight task. Furthermore, pupil dilation has also been shown to increase with an increase in autonomic arousal, and with increasing cognitive workload (Bradley et al., 2008; Einhäuser et al., 2008; Hyönä et al., 1995; Kahneman, 1973; Marinescu et al., 2018; Marshall, 2002). Therefore, when comparing the physiological data from other research, the results from the current study indicate that in response to an unexpected engine failure, pilots show arousal consistent with the startle reaction or high stress. Furthermore, pilots appear to have arousal analogous to high workload in the expected engine failure.

Consistent with the hypothesis, during the unexpected stall, pilots had a significantly larger pupil dilation compared to before the event. In contrast, there was no significant difference in pupil dilation during the expected stall event, compared to before the event. Contrary to the hypothesis, the increase in mean increase heart rate compared to the baseline flight was significantly larger for the expected stall, compared to the unexpected stall. There were no significant differences in mean heart rate between the unexpected and expected stall events and the baseline flight. These data are conflicting as the pupil dilation shows a significant increase in arousal in the unexpected stall only, while the heart rate data indicates a larger change in heart rate from baseline in the expected stall compared to the unexpected stall. Both increases in pupil dilation and heart rate can indicate an increase in cognitive workload and autonomic arousal (Bradley et al., 2008; Einhäuser et al., 2008; Hyönä et al., 1995; Kahneman, 1973; Marinescu et al., 2018; Marshall, 2002). On account of this, it is concluded that both stall types appear to have led to physiological arousal in the pilots.

To arrange an unexpected stall independent of the pilots' actions required the simulator to be programmed with major changes in wind speed and direction. The coding of the aerodynamic stall only led to just over 50% of pilots encountering the stall horn. The pilots that did not encounter the stall horn were excluded from the stall analyses. Therefore, for the stall flights the sample size was quite small ($N = 13$). Due to this the power of the statistical tests were limited. Therefore, some of the following trends that are described were based on data from non-significant analyses. The non-significant data suggest that pilots had a higher heart rate in the expected stall event, compared to the unexpected stall event. Additionally, both stall events elicited higher heart rates than the baseline flight.

Pilots showed a similar increase heart rate for the expected stall (4.93bpm) as the expected engine failure. This heart rate increase value (around 4-5bpm) is thought to be consistent with increases in heart rate in response to an increased cognitive workload, as shown by previous research (Fallahi et al., 2016; Grassmann et al., 2017). Compared to the expected stall, in the unexpected stall the mean increase in heart rate from the baseline flight value was small (less than 1.5bpm). Fallahi et al. (2016) found that workers from cement, city traffic control and power plant control had an increase in heart rate of approximately 2bpm during low mental workload tasks in comparison to baseline. In relation to this finding, the heart rate increase in the unexpected stall may be due to arousal from a small increase in mental workload. Neither stalls showed physiological arousal consistent with the startle response (Chou et al., 2014; Deuter et al., 2012; Holand et al., 1999). Due to the very different nature of the two stall events as well as the contradicting pupil dilation and heart rate findings, it is not possible to make comparative conclusions between the expected and unexpected stall.

Inducing startle or high levels of stress via unexpected events may rely on the type of abnormal flight event encountered and the subjective appraisal of its severity. It is possible that the unexpected engine failure led to a higher increase in heart rate than the unexpected stall due to the pilots realizing that they would be required to land. In contrast, the unexpected stall only required an adjustment to plane handling. Previous research has shown that subjective appraisals of threat have an impact on the physiological stress response. Hodges and Spielberger (1966) investigated the effects of severity of fear (of electric shock) on participants' heart rates. Participants were required to perform a pursuit-rotor task, this involved following a small circle on a rotating turntable. After a few practices, some subjects were threatened with shock if they performed poorly, others were threatened with shock regardless of performance and some participants were not threatened. No shocks were actually given, however subjects with a high fear of shock subjects reacted to the threat of inevitable shock with significantly greater increases in heart rate than low fear of shock subjects. Therefore, the results of Hodges and Spielberger (1966) study as well as the present study support the intuitive idea that the consequences of each event impact the magnitude of the physiological response. These findings are also consistent with Landman's (2017a) conceptual model, where the perceived intensity of the event can lead to startle (Figure 1, repeated below). Pilots may have had a higher fear of emergency landing than stall. As a stall that is encountered and safely recovered from is not an emergency. This may explain the larger physiological responses in the engine failure events compared to the stall events.

The physiological changes found in response to an unexpected engine failure in the simulator are a positive finding in terms of its implications for training. It appears that unexpected flight events, even without a concordant loud noise, or preceding distraction can lead to physiological response analogous to startle. In her article on the technical difficulties of simulating the element of surprise in upset recovery training; Bürki-Cohen (2010) stated that it may be impossible to generate the exact physiological response in the safety of a simulator. However the current research shows that it is possible to create a physiological response analogous to startle in a fixed- base simulator. Bürki-Cohen (2010) and Martin et al. (2015) suggest creating an in-flight atmosphere conducive to startle; stressing pilots with realistic tasks and distractions, or placing a loud auditory stimulus before the startling or surprising event. This is to moderate the "simulator mindset" and offset the absence of real-life risk. Although these suggestions would be helpful for pilot training, they may be confounding in quantitative research on startle. The current research shows that these confounding stimuli are not necessarily required and some events will lead to arousal consistent with startle simply when they occur unexpectedly and in conditions analogous to real life.

**Information Processing**

When an unexpected flight event occurs, pilots need to assess and diagnose the situation. Therefore, it was hypothesized that pilots would spend more time looking inside at the cockpit flight display screen and key instruments during unexpected events compared to expected events. Contrary to the hypothesis, during the unexpected engine failure pilots spent significantly less time overall looking at the MFD and PFD displayed on the LCD monitor compared to the expected engine failure. Furthermore, contrary to the hypothesis during the unexpected engine failure, the pilots spent significantly less time looking at the airspeed indicator, the altimeter, the GPS, the turn coordinator, and the directional gyro compared to the expected engine failure. There were no significant difference in the time spent viewing the attitude indicator, the vertical speed gauge, the navigational instruments or the engine instruments.

Consistent with the hypothesis pilots spent significantly more time viewing inside the cockpit (head down display) at the MFD and the PFD during the unexpected stall compared to the expected stall. Specifically, pilots spent significantly more time looking at the airspeed indicator, the attitude indicator, the GPS, the turn coordinator, and the engine instruments in the unexpected stall. There were no significant differences in the amount of time pilots spent time viewing the vertical speed indicator, the directional gyro and the altimeter in the unexpected stall compared to the expected stall.

Pilots are taught that during an engine failure they should rely on external cues instead of relying solely on the internal controls. This is reflected in the fact that during both engine failures the pilots spent more than 90% of their time viewing outside of the cockpit. However, the significant difference between the amount of time spent viewing the controls during the expected engine failure ($M = 6.51\%$) compared to the unexpected engine failure ($M = 3.73\%$) may reflect a harmful decrease in viewing the flight displays. This may suggest that pilots are not spending enough time viewing crucial instruments which will help them to land safely during the unexpected engine failure.

The investigation into the percentage time spent by the pilots looking at the different controls is exploratory research. No specific hypotheses were made regarding this data. It is also recognized that multiple comparisons are occurring. The significance level was adjusted so that any analyses where $p > .025$ were not considered significant. This was done to prevent against Type I errors, as well as Type II errors that occur with a full Bonferroni correction. However it is recognized that this correction may not have been conservative enough.

A reduction in the observation of the critical instruments during the unexpected engine failure could be attributed to attentional tunnelling. Visual and cognitive processing are thought to occur during fixations (Salvucci & Goldberg, 2000). The physiological data suggest that the unexpected engine failure led to more autonomic arousal than the expected engine failure. Therefore, the unexpected engine failure was likely more stressful than the expected engine failure. Stress has also been shown to lead to a reduction in peripheral cue utilisation and narrowing of the attentional field (Baddeley, 1972; Combs & Taylor, 1952; Easterbrook, 1959; Staal, 2004). Furthermore, Streufert and Streufert (1981) concluded from their research that danger affects performance by impacting an individual's breadth of attention and ability to logically and calmly process information. In the present study, fixations are the percentage time participants spent looking at the individual AOI. Pilots spent significantly less time observing the airspeed indicator, altitude meter, GPS, and directional gyro, in the unexpected engine failure flight compared to the expected engine failure. The data in the present study could indicate attentional tunnelling in response to unexpectedness or high arousal; where attention is focused on the threat, possibly the external environment in this case, and less concentration on peripheral cues, in this case the flight display instruments.

Disordered attentional processing due to unusually high levels of arousal and vigilance has been shown to result in an indiscriminate search, fast disordered attentional shifting and a reduction in the number and quality of other courses of action considered (Staal, 2004). The decrease in percentage time spent viewing the flight display during the unexpected engine failure may be suggestive of fast disordered attentional shifting. In which information processing was disrupted and the pilots shifted their attention so quickly that the Gazepoint software did not recognize some fixations. The Gazepoint samples at 60Hz, which is a relatively low modern sampling rate (Raney, Campbell, & Bovee, 2014). A sampling rate of 60 Hz samples the eye position every 16.7 msec (Raney, Campbell, & Bovee, 2014). A 60Hz sampling rate will lead to an average error of approximately 8 msec, which could be considered too large to study the duration of saccades, but not too large to study the duration of fixations (Raney, Campbell, & Bovee, 2014). Thus, pilots decrease in fixations during the unexpected engine failure may indicate fast disordered attentional shifting in the form of brief saccades. The eye-tracking data therefore support and extend a previous plethora of research on attentional tunnelling and information processing disruption under stress in an operational context.

Pilots spent almost four times as long looking at the flight display screen in the unexpected stall compared to the expected stall. Pilots spent more of their time looking at most of the instruments during the unexpected stall, even unrelated instruments such as the GPS and the engine instruments. Therefore, pilots appear to be attempting to gather and process as much information as possible. This is likely to indicate information search strategies aiming at assisting situational diagnosis. Previous research has found that less experienced operators spend a longer amount of time looking at displays, as well as looking at a larger percentage of the information (Kirschenbaum, 1992; Wiggins & O'Hare, 1995). The search strategy seen in the unexpected stall is similar to the search strategies used by less experienced operators.

Only the pilots that had a stall horn sound following the decrease in headwind were included in the analysis. It would be expected that training and experience would lead the stall horn to be a salient indicator of a stall. However, these eye-tracking findings suggest that the stall horn may not automatically initiate the required aerodynamic stall schema or frame (Landman et al., 2017a) (Figure 1, repeated below). Furthermore, during the experiments, it was common for the pilots to express confusion about the event after the unexpected stall flight. Therefore, the practice stall template may not be creating automatic associations with the stall horn because the practice stall is too different than real-life windshears and stalls. This may lead to pilots having to spend a long time uncertain of the flight situation, and trying to diagnose it. This is dangerous as a stall requires immediate recovery response, especially at low altitudes.

**Performance**

When comparing performance data, it is almost certain that the analyses were limited by low power due to small sample sizes. Therefore this part of the study was mostly exploratory in nature, and therefore the findings and conclusions should be viewed and considered as such. Perhaps these findings will inspire future research.

It was hypothesized that in the unexpected events pilots would perform more poorly compared to the expected events (e.g. more crashes in engine failure, more altitude lost in the stall, and more incorrect responses). Concordant with this in the unexpected engine failure approximately 45% of the pilots crashed, where no pilots crashed in the expected engine failure. Contrary to the hypotheses, the same percentage of pilots in both engine failure flights forgot to pull back the throttle after the engine had failed to prevent engine surges affecting the plane's glide. Concordant with the hypothesis, the remaining participants showed a slower response time to pull back the throttle in the unexpected engine failure compared to the expected engine failure. However, this difference was not significant.

The altitude lost and the time to recover in the stall conditions were not comparable between the expected and unexpected stall flights as the configuration of the two flights were too different. In the unexpected stall flight, pilots already had maximum power before the stall occurred. However, concordant with the hypothesis during the unexpected stall approximately 31% of pilots incorrectly pulled back on the throttle in response to the stall horn, compared to zero percent in the expected stall. As well as this, in the unexpected stall 38% of pilots did not lower the plane nose in response to the stall horn, compared to zero percent in the expected stall. This indicates that in unexpected events pilots can show impaired performance whereas in the expected stall events pilot's performance was optimal across the board. This may be due to pilots not recognizing the unexpected stall, lack of recognition combined with an incorrect response to the stall horn at 500ft could prove a lethal combination in a real life setting. Consequently, such a high percentage of incorrect responses is alarming.

Pilots showed poorer performance in unexpected events compared to expected events. The poorer performance in the unexpected events is consistent with aviation safety reports and previous research. In a similar study, Casner et al. (2012) showed that pilots' performance is inferior in unexpected events, compared to when they are expected. As well as this aviation accident reports show that pilots are sometimes responding inappropriately to what should be well-practiced emergency events, for example not directing the plane downwards following a stall warning (NTSB; 1995, 2004, 2010a, 2010b). This research also supports findings from Ledegang and Groen (2015) who reported that pilots have trouble recovering from aerodynamic stalls when they have not recently reviewed the procedure. As pilots in the present study had not always reviewed the stall recovery procedure before encountering the unexpected stall. Furthermore, the findings support Landman et al. (2017b) and Schroeder et al. (2014) who also found that in response to an unexpected stall a large percentage of pilots did not adhere to the standard operating procedures. The current research combined with previous findings, and accident reports suggest that pilot performance suffers when abnormal events are encountered in an unexpected manner.

**Performance and Arousal**

It was hypothesized that pilots that show impaired performance would show high arousal consistent with the startle reflex. Contrary to this hypothesis there were no significant differences in heart rate or pupil dilation between those participants that crashed and those that landed safely. Contrary to the hypothesis there were no significant differences in the heart rates or pupil dilations of pilots that responded correctly and pilots that responded incorrectly following the stall horn.

The following discussion is speculation on non-significant data trends. When looking at the non-significant values there are both trends that support and do not support the hypotheses that pilots that show impaired performance would show high arousal consistent with the startle reflex. Contrary to the hypotheses pilots who crashed in the unexpected engine failure had lower mean heart rates during the event compared to the individuals that landed safely. Concordant with the hypothesis pilots who correctly kept the throttle at full power and pilots who correctly directed the nose downwards had lower heart rates than pilots that implemented incorrect responses to the stall horn. Furthermore, pilots that incorrectly pulled back on the throttle had a larger pupil dilation but a smaller change in pupil dilation compared to pilots that exhibited the correct response. Consistent with the hypothesis, pilots that failed to pitch the nose downwards following the stall horn had a larger pupil dilation and a larger increase in pupil dilation compared to the pilots that responded correctly. Thus, there is some preliminary evidence supporting the idea that increased arousal may lead to incorrect responses to the unexpected stall, however due to lack of power and small sample size this conclusion is tentative at best. Future research should investigate this relationship, and its causality.

Contrary to what was expected, this present study did not find that pilots that performed poorly had high autonomic arousal analogous to startle. Research on fear conditioning has shown that when startle occurs in the presence of perceived threat the response can become exacerbated leading to what is known as fear potentiated startle (Bradley, Moulder, Lang; 2005). The fear potentiated startle that occurs due to a combination of a startling stimulus, and the perception of threat can lead to a fully developed stress reaction (Eysenck, Payne, & Derakshan, 2005). Research shows that this stress can impair information processing (Eysenck et al., 2005; Thackray & Touchstone, 1983), as well as reduce working memory capacity (Bradley et al., 2005). An increase in heart rate and pupil dilation in the simulated unexpected engine failure in the present study, supports the well-accepted idea that startle due to an unexpected flight emergency may lead to fear potentiated startle and consequently impaired decision making and behaviour. The presence of the arousal analogous to startle is suggestive that in the presence of real threat, the startle could develop into to a fully-fledged fear potentiated startle response.

**Performance and Information Processing**

It was hypothesized pilots that crash in the unexpected engine failure would spend more time looking at the displays than pilots that landed safely, as the pilots may forget to fly the plane and focus on the controls. Contrary to the hypotheses there were no significant differences in the amount of time pilots spent looking at the flight display (MFD and PFD) between pilots that crashed and pilots that landed safely. It was also hypothesized that pilots that performed better on the unexpected engine failure would have similar information gathering strategies as revealed by the eye tracker data. Contrary to the hypothesis there were no significant differences in the amount of time pilots that crashed and pilots that landed safely spent looking at the individual flight instruments.

This section was limited by small sample sizes due to between-subjects comparisons. It is thought that these statistical analyses were limited by lower power due to these small sample sizes. For example, thirteen pilots incorrectly did not lower the nose of the plane and pitch it downwards in response to the stall horn. However, the non-significant findings are interesting and consistent with aforementioned findings and theories, therefore the following examination of the nonsignificant trends was reported. However, these observations are tentative and only indicate future research possibilities instead of conclusive findings.

Following an engine failure in a small aircraft under visual flight rules, pilots are taught to focus on the environment outside the aircraft, with some reliance on the airspeed indicator and the altimeter. The airspeed is used to stay on the best glide speed and ensure a safe landing speed, and the altimeter is used for constantly checking height above ground when working out where to land. A quick scan of other instruments is also useful to watch for any change in aircraft state, for example partial power may return. Pilots that landed safely spent more time observing the airspeed indicator and the altimeter compared to pilots that crashed, although these differences were not significant. Pilots that crashed spent more time focusing on the navigational instruments, and the GPS. There were no substantial differences for the remaining AOIS or instruments. Consequently, it appears that pilots that perform well are better able to simplify their focus to the important instruments which will help them to retain control of the aircraft and operate in a safe manner.

When looking at the non-significant differences pilots that landed safely spent more time viewing the airspeed indicator, the attitude indicator, and the altimeter. Whereas, pilots who crashed spent more time viewing the navigational instruments and the GPS. Pilots that crashed and pilots that landed safely spent a similar percentage of time viewing the turn coordinator, the engine instruments, the directional gyro, and vertical speed indicator. Interestingly when comparing Figure 14 and Figure 16, pilots that landed safely had similar information search strategy as pilots did in the expected engine failure. Both pilots that landed safely in the unexpected engine failure, and pilots overall in the expected engine failure, spent more time observing the airspeed indicator, the attitude indicator, and the altimeter. These findings suggest that the unexpectedness of the engine failure disrupts the effectiveness and efficiency of some pilots search strategies.

Pilots that crashed appear to be spending longer looking at instruments that are informative about their location. Pilots are taught that while flying they should be observant of possible landing places in case of an engine failure. In the simulated flight it was not possible to look outside the side windows of the cockpit. Therefore, after the occurrence of the engine failure, pilots were required to rely on their memory of the terrain they had subsequently passed, or pick a landing point in front of them. It is possible the pilots that crashed had a lack of situational awareness, were unsure of where to land, and therefore spent more time observing the navigational instruments and the GPS. This time may have been better spent attempting to safely landing the plane in an area in their direct line of sight.

It was also hypothesized that pilots that performed better during the unexpected stall would have similar information gathering strategies as revealed by the eye tracker data. Pilots that incorrectly did not point the nose of the plane down spent significantly less time viewing the altimeter, and the vertical speed indicator. When looking at the nonsignificant differences, pilots that did not pitch the nose of the plane downwards, spent less time viewing inside the flight displays than the pilots that responded correctly. Pilots that incorrectly did not pitch the nose of the plane down spent more time looking at the GPS, and less time viewing the turn coordinator.

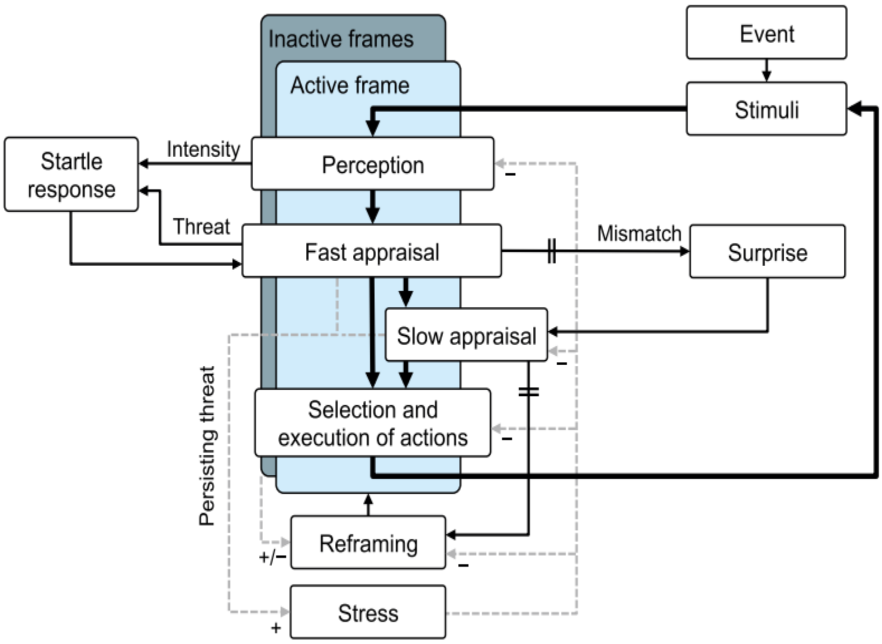
Contrary to the hypothesis there were no significant differences when looking at the information search strategies for pilots who incorrectly pulled back on the throttle following the stall horn. However, these investigations are almost certainly limited by low power as only four of thirteen pilots incorrectly pulled back on the throttle after hearing the stall warning. Therefore, again, the following data trends, can only suggest tentative conclusions. Pilots that incorrectly pulled back on the throttle following the stall horn, spent longer viewing the flight displays. Pilots that incorrectly pulled back on the throttle following the stall horn spent less time looking at the airspeed indicator, and the turn coordinator.

Collectively the data support the hypothesis and indicate that there may be differences in the information processing strategies between pilots that perform poorly in unexpected abnormal events. As even with a low powered analysis it was found that pilots that incorrectly did not point the nose of the plane down spent significantly less time viewing the altimeter, and the vertical speed indicator. It is possible that with a higher powered analysis other non-significant differences mentioned might also be confirmed differences.

The pilots that performed poorly may not have recognized the unexpected stall. This is proposed as the pilots that did not pitch the nose down spent significantly less time viewing the altimeter and the vertical speed indicator, which are the two instruments that are important and helpful in stall recovery. Although non-significant, there was a trend that the pilots that did not pitch the nose down spent a larger percentage time viewing the GPS. The GPS is a goal orientated instrument displaying the flight path to Stewart Island (the landing destination). This suggests that the pilots that did not pitch the nose down, may have just continued to fly towards their destination not recognizing the presence of the abnormal stall event. Additionally pilots that did not pitch the plane down, spent less overall time viewing the flight display. This may indicate pilots that did pitch the nose of the plane down spent more time diagnosing the abnormal stall. In contrast pilots that incorrectly pulled back on the throttle recognized that there was an abnormal event but initially diagnosed it incorrectly, as they implemented a response directly opposite to the trained stall recovery response. This is also indicated by the fact that they spent longer looking at the flight display (PFD and MFD), which may indicate a longer percentage time spent diagnosing the situation.

**Landman et al's (2017a) conceptual model of startle and surprise**

Pilots physiological responses to the unexpected engine failure in the current research supports Landman et al. (2017b), indicating that surprise and startle can be produced in simulated environments following an unexpected event and can therefore be implemented to help train upset recovery responses. When looking at the results of this study in terms of Landman et al's (2017a) conceptual model, the findings suggest that in the engine failure flights the unexpected event was more startling than the expected event due to its intensity or perceived threat (Figure 1). According to Landman et al's (2017a) conceptual model of startle and surprise in terms of sense-making and decision making, in the unexpected event the pilots will likely be operating in a normal flight frame or schema (Figure 1).When the abnormal flight event occurs, there will be a frame mismatch and pilots will have to switch to a recovery schema specific to the abnormal event which they would have learnt throughout their training (Landman et al., 2017a). However, the finding that pilots spent less time looking at the flight instruments during the unexpected engine failure does not support this idea, and indicates fast appraisal and selection of response (Figure 1). This may be due to emergency landings having the same principles as normal landing, therefore the selection and execution of actions can be completed within the active frame of normal flight and require no frame switch. However it may also indicate, that because an engine failure is very salient and well-practiced, the frame switch may be automatic.




*DUPLICATE Figure 1.* Conceptual model of startle and surprise. Solid lines indicate sequenced events. Dashed lines indicate potential influences, with plus signs indicating an increasing effect and minus signs indicating an impaired effect. Double lines indicate thresholds. Reprinted from “Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise” by A Landman, E Groen, MM van Paassen, AW Bronkhorst, and M Mulder, 2017, Human Factors, 59, p. 1163. Copyright 2017, by the Human Factors and Ergonomics Society.

According to Landman et al’s (2017a) conceptual model the results from this study suggest that the frame switch for the unexpected stall, was less efficient or less accurate compared to the unexpected engine failure. Pilots spent a significantly larger percentage of time looking at the flight displays (PFD and MFD) during the unexpected stall event, compared to the expected stall. According to the model this may indicate that there was a mismatch after fast appraisal leading to surprise and then slow appraisal (Figure 1). Some pilots incorrectly pulled back on the throttle following the stall horn, which indicates that the mismatch occurred after the first selection of actions (Figure 1). According to the model the pilots would then start back again at the stimuli perception stage (Figure 1). This is supported by the nonsignificant trend that pilots who incorrectly pulled back on the throttle spent longer looking at the flight instruments than those that responded correctly.

When looking at other nonsignificant differences, pilots that did not pitch the nose of the plane downwards spent less time viewing the flight display (PFD and MFD) than pilots that correctly responded. In terms of Landman et al’s (2017a) model, these pilots did not recognise the cues or stimuli, and therefore incorrectly remained in the same frame of normal flying (Figure 1).The lack of physiological arousal in the unexpected stall, indicates that although the pilots may have been surprised, they were not startled, which according to the model is due to lack of perceived threat or intensity (Figure 1). Although overall pilots had lower arousal levels in the unexpected stall compared to the expected stall, the results showed that pilots that responded incorrectly to the stall had higher heart rates, and larger pupil dilations compared to those who implemented the correct recovery. These findings are concordant with Landman et al’s (2017a) theory that stress leads to issues with reframing such as choosing an incorrect frame, confusion, or loss of situational awareness. Further research could investigate the differential responses to abnormal events in accordance to Landman et al. (2017a) model.

The findings of the current research support McKinney Jr and Davis's (2003) conclusions that familiarity due to practice impacts performance. McKinney Jr and Davis (2003) defined a wholly practiced event as a scenario including a malfunction that pilots had deliberately practiced. Whereas a partially practiced event was one where the malfunction occurred within a wider emergency situation which was novel, and had not been practiced. The unexpected stall was induced in a manner that pilots would not likely have experienced before. Furthermore, after the unexpected stall flight many pilots indicated that that type of event is very rare in New Zealand. Concordant with this, a report by the FAA suggests that windshear encounters occur infrequently (FAA, 1988). Therefore the unexpected stall condition induced by sudden change of wind speed direction would likely fall into McKinney Jr and Davis's (2003) category of 'partially practiced'. In support of McKinney Jr and Davis's (2003) findings, this present research found impaired responses to an unexpected and unfamiliar stall. This calls into question the practice stall method used in normal general aviation training as pilots induce the stall themselves. As this would not likely occur in real life, stalls occurring during actual flight would always be 'partially practiced'.

Flight training occurs in ideal conditions (i.e. perfect landing terrain, and high altitude) which is not likely to always be comparable to real life flight emergencies. In the unexpected engine failure there were a number of potential areas to land in, however the terrain was hazardous and uncertain. Furthermore, in engine failure training pilots are not required to land (CAA, 2012) . Differences between training engine failures and the unexpected engine failure that occurred in this study, would define the unexpected engine failure as partially practiced as well. Performance was impaired in the unexpected versus the expected flight conditions, which is consistent with the findings of McKinney Jr and Davis (2003) where performance was not improved for event that were only partially practiced.

Therefore, training pilots for abnormal events in ideal conditions may not help them deal with actual real-life emergencies which require quick thinking and flexibility.

**Conclusion**

Consistent with the hypotheses both expected and unexpected versions of abnormal flight events lead to increases in physiological arousal in pilots. Furthermore, the unexpected engine failure lead to arousal similar to startle, however it was not possible to make comparative conclusions about arousal in the expected and unexpected stall. These differences were hypothesized to be due to different subjective appraisal of severity, intensity or threat. Results from this study show that it is possible to induce arousal indicative of startle in a flight simulator. This has positive implications in terms of the possibility of using simulators to train for unexpected abnormal events, which may mediate the negative cognitive effects of startle. Pilots have differential information processing strategies for unexpected and partially practiced events compared with expected and wholly practiced events. This may be due to attentional tunnelling or lack of recognition, and may indicate that pilots require more training in controlling the information search strategies upon encountering abnormal flight events. The current research combined with previous findings, and accident reports suggest that pilot performance suffers when abnormal events are encountered in an unexpected manner. This may be due to lack of recognition and ability to adapt training to novel situations. The results of this study extend the evidence supporting Landman et al’s (2017a) recent conceptual model of startle and surprise. Furthermore, there are many indications of areas that could be further investigated, e.g. characterising the different main incorrect responses to abnormal events, in terms of Landman et al’s (2017a) study, which may help to inform training templates.

# Chapter 4: Study 3

## Overview

*Method*

This study used the combined data from Studies 1 and 2 to investigate hypotheses related to expertise and performance. Data from Study 1 were used as a 'novice' comparison, and pilots were separated into 'intermediate' (student and privately licenced) and 'expert' (commercially licenced) pilots. This third phase of the research compared the data over the three difference experience levels.

*Hypotheses*

It was hypothesized that during the engine failure and stall flight events, expert pilots would have the lowest heart rate and smallest heart rate difference from baseline, compared to intermediate pilots and novices, with novices having the highest. It was hypothesized that during the engine failure and stall flight events, expert pilots would have the smallest pupil dilation and change in pupil dilation compared to intermediate pilots and novices, with novices having the largest. It was hypothesized that expert pilots would spend less time looking inside the cockpit at the MFD and PFD, than the intermediate pilots during the unexpected events, because they would be faster at diagnosing the situation and quicker to implement an effective information search protocol. It was hypothesized that intermediate pilots would have the greatest amount of time spent looking inside the cockpit (head down display) at the MFD and PFD, as Wiggins and O'Hare (1995) found intermediate pilots appear to try and gather as much information as possible. It was hypothesized that novices would spend the least amount time viewing the flight display (MFD and PFD) during the unexpected stall, as it is likely that they will not identify that there is an abnormal flight event. It was hypothesized that novices will spend the most 'head down' time looking at the flight displays in the unexpected engine failure as they will need to spend more time diagnosing the situation and they had not been taught to focus on the external environment during engine failures. It was hypothesized that expert pilots would perform the best compared to intermediate pilots and university students, and intermediate pilots would perform better than university students. Furthermore, it was hypothesized that expert's search strategies would be more efficient in regards to each specific event, compared to the intermediate pilots or the university students.

## Method

The Survey Monkey questionnaire which pilots filled out before the simulator tasks included questions regarding pilot's personal information and pilot experience (Appendix G). Information collected from this included: pilot's total flight hours, and pilot's recent flight hours, and pilot's licence. Participants were separated into three groups according to their licence; University students (no licence), Student and private licenced (intermediate) pilots, and commercial licenced (experts) pilots. Physiological response, information search strategy, and performance data were compared over the three different levels of experience.

## Results

### Data Screening

The data screening process was the same as Study 1.

### Heart Rate

The design of the of the following analyses was a 2 way mixed ANOVA where Expertise had three between subject levels (university students, intermediate pilots, and expert pilots) and flight type had three within subject levels (baseline, expected event, and unexpected event).

Pilot 15 was excluded from heart rate analysis as their heart rate values were abnormal (see Study 2 for further information).

***Engine Failure***

*Mean Heart Rate*

The sample sizes for the engine failure flights were as follows; University students ($N$ = 21), intermediate pilots ($N$ = 12), and expert pilots ($N$ = 9). A two-way mixed ANOVA was completed to assess the effect of expertise on mean heart rate during the different flight events.

University students had a similar heart rates in the unexpected engine failure ($M$ = 83.01, $SD$ = 13.70), and the expected engine failure ($M$ = 82.96, $SD$ = 13.10), which were in turn slightly higher than the baseline heart rate ($M$ = 81.46, $SD$ = 12.16) (Figure 17). Intermediate pilots had a higher heart rate in the unexpected engine failure ($M$ = 100.44, $SD$ = 15.09), compared to the expected engine failure ($M$ = 95.66, $SD$ = 16.56)**,** which was in turn higher than the baseline heart rate ($M$ = 91.65, $SD$ = 11.60) (Figure 17). Expert pilots had a higher heart rate in the unexpected engine failure ($M$ = 95.32, $SD$ = 14.59), compared to the expected engine failure ($M$ = 91.19, $SD$ = 15.50), which was in turn higher than the baseline heart rate ($M$ = 86.03, $SD$ = 14.51) (Figure 17).

The presence of outliers in the different heart rate data for each flight was examined over each level of experience using boxplots. There was one outlier in the expected engine flight for intermediate pilots, the outlier was assessed and was considered plausible, and therefore it was kept in the analysis. There were no outliers in the unexpected engine failure and baseline flights over the three different expertise levels. The data were normally distributed for each flight type over each of the three levels of expertise (Shapiro-Wilk, $p > .05$). There was homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi2(2) = 4.222$, $p = .121$.

There was a statistically significant interaction between the flight type and experience level on mean heart rate, $F(4, 78) = 3.8633$, $p = .009$, partial η2 = .157. There were no statistically significant differences in mean heart rate (BPM) between experience levels during the baseline flight. There was no statistically significant difference in the mean heart rate (BPM) between experience levels during the expected engine failure event. There was a statistically significant difference in mean heart rate (BPM) over the three different experience levels during the unexpected engine failure event, $F(2, 39) = 6.300$, $p = .004$, partial η2 = .244.

Data reported are mean differences between groups. Bonferroni post hoc tests indicated that heart rate was significantly higher in the intermediate pilot group ($M = +17.43$, $SE = 5.17$, $p = .005$), compared to the university students (Figure 17). Expert pilots had a lower heart rate than intermediate pilots ($M = -5.12$, $SE = 6.30$, $p = .698$), but this was not significant (Figure 17). Expert pilots had higher mean heart rates than the university student group ($M = +12.31$, $SE = 5.69$, $p = .091$), however this was not statistically significant (Figure 17).

*Change in Heart Rate*

University students had the smallest change in heart rate during the expected engine failure compared with baseline ($M = 1.50$, $SD = 4.29$). The university students were followed by the intermediate pilots ($M = 4.01$, $SD = 7.18$), who had a smaller change in heart rate than the expert pilots ($M = 5.16$, $SD = 4.43$). During the unexpected engine failure university students had the smallest increase in heart rate compared to baseline ($M = 1.55$, $SD = 5.77$), followed by the intermediate pilots ($M = 8.80$, $SD = 7.61$), who had a slightly smaller increase in heart rate than the expert pilots ($M = 9.29$, $SD = 8.92$). A two-way mixed ANOVA was completed to assess the effect of expertise on change in mean heart rate compared to baseline over the expected and unexpected engine failures.

Due to outliers and homogeneity of variance being violated the data was transformed via a square root transformation. The presence of outliers in the change in heart rate data for each flight over each level of experience was examined using boxplots. There were no outliers in the unexpected engine failure flight or the expected engine failure for the intermediate and expert pilots. Also there was no outliers in the expected engine failure for the university students. There was one outlier in the unexpected engine flight for university students. The data were normally distributed for each flight type over each of the three levels of expertise (Shapiro-Wilk, $p > .05$). The outlier in the university student group was assessed and was considered erroneous and was therefore removed from the analysis. There was homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively.

The two way mixed ANOVA indicated that there was no significant interaction effect between flight type and expertise ($p = .466$). There was no significant main effect of flight type on change in heart rate ($p = .255$). However, there was a significant main effect of expertise level on change in heart rate $F(2, 20) = 11.216$, $p = .001$, partial η2 = .529. Post hoc analysis with a Bonferroni adjustment revealed that university students had a mean change in heart rate 1.17bpm, 95% CI [0.38, 1.95] lower than intermediate pilots, a statistically significant difference $p$ =.003. As well as this university students had a mean change in heart rate 1.47bpm, 95% CI [0.38, 1.95] lower than expert pilots, a statistically significant difference $p$ =.002. There was no statistically significant differences in change in heart rate between the expert and the intermediate pilots.

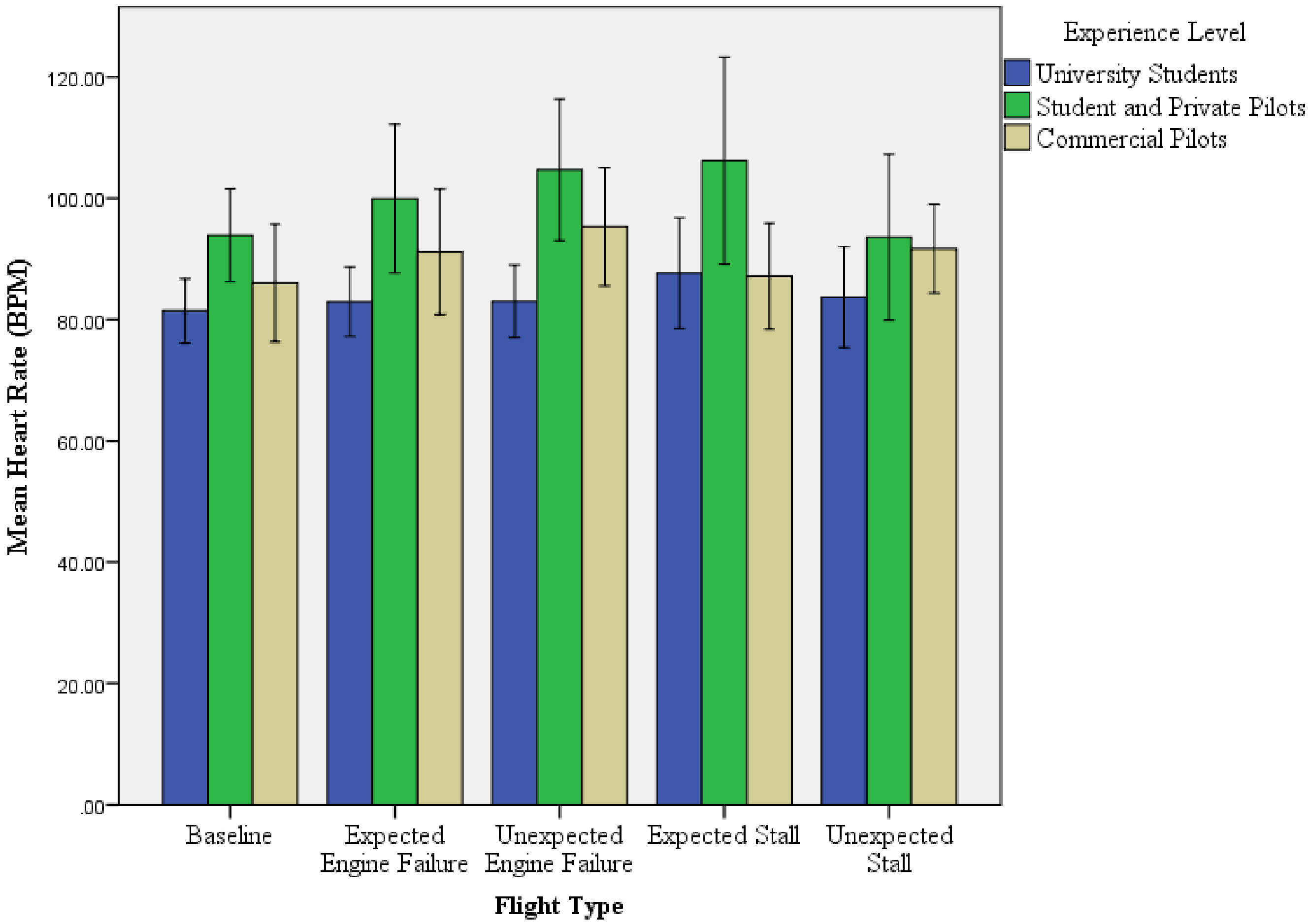



*Figure 17.* Mean heart rate (BPM) for university students, student and private (intermediate) pilots, and commercial (expert) pilots for each experimental flight with standard error (+/- 2SE).

***Aerodynamic Stall***

*Mean Heart Rate*

Sample sizes for the stall flights were as follows; University students (No licence) (*N* = 9), Student and private licenced (intermediate) pilots (*N* = 8), and commercial licenced (expert) pilots (*N* = 8). A two-way mixed ANOVA was completed to assess the effect of expertise on mean heart rate during the different flight events.

University students had a higher heart rate during the expected aerodynamic stall (*M* = 87.68, *SD* = 13.69), compared to the unexpected stall (*M* = 83.71, *SD* = 12.48), which was in turn lower than the baseline flight heart rate (*M* = 86.06, *SD* = 13.15) (Figure 17). Intermediate pilots had a higher heart rate during the expected aerodynamic stall (*M* = 99.17, *SD* = 15.41), compared to the unexpected stall (*M* = 89.68, *SD* = 10.48), which was also lower than the baseline heart rate (*M* = 92.12, *SD* = 11.19) (Figure 17). Commercial pilots had a higher heart rate during the unexpected aerodynamic stall (*M* = 86.47, *SD* = 10.83), compared to the expected stall (*M* = 83.71, *SD* = 12.48), both stall events elicited a higher heart rate from the expert pilots than in the baseline flight (*M* = 84.37, *SD* = 14.56) (Figure 17). A two-way mixed ANOVA indicated that there was no statistically significant interaction between the flight type and experience level on mean heart rate ($p$ = .066). There was no significant main effect of flight type on mean heart rate ($p$ = .069). There was also no significant main effect of expertise level on mean heart rate ($p$ = .094).

*Change in Heart Rate*

University students had the smallest change in heart rate compared to the baseline flight during the expected stall (*M* =1.62, *SD* = 4.49). The expert pilots also had a small change in heart rate in the expected stall (*M* = 2.80, *SD* = 6.27), compared to the intermediate pilots who had the largest change in heart rate compared to the baseline flight (*M* = 7.05, *SD* = 7.36). During the unexpected stall the expert pilots had a larger mean increase in heart rate (*M* = 5.32, *SD* = 15.70), whereas both the intermediate pilots (*M* = -2.78, *SD* = 4.86), and the university students (*M* = -2.36, *SD* = 4.65) had a decrease in heart rate compared to the baseline flight. The intermediate pilots had a larger decrease in heart rate than the university students.

A two-way mixed ANOVA was completed to assess the effect of expertise on change in mean heart rate compared to baseline over the expected and unexpected stalls. The analysis indicated that there was no significant interaction between flight type and expertise ($p$ = .063). There was no significant main effect of flight type on change in heart rate ($p$ = .378). Additionally there was no significant main effect of expertise on change in heart rate ($p$ = .069).

**Pupil Dilation**

The design of the of the following analyses was a 2 way mixed ANOVA where Expertise had three between-subject levels (university students, intermediate pilots, and expert pilots) and flight type had two within-subject levels (expected event, and unexpected event).

***Engine Failure***

*Mean Pupil Dilation*

University students had the largest pupil dilation ($N$= 18, $M$ = 19.55, $SD$ = 2.11) in the expected engine failure, followed by the commercial pilots ($N$= 6, $M$ = 19.37, $SD$ = 3.78), and then the student and private licenced pilots ($N$= 11, $M$ = 18.11, $SD$ = 2.85) who had the smallest pupil dilation (Figure 18). University students had the largest pupil dilation ($N$= 18, $M$ = 18.75, $SD$ = 2.12) in the unexpected engine failure, followed by the commercial pilots ($N$= 6, $M$ = 18.70, $SD$ = 3.70), and then the student and private licenced pilots ($N$= 11, $M$ = 17.22, $SD$ = 3.19) who had the smallest pupil dilation (Figure 18).

The presence of outliers in the different mean pupil dilation data for each flight over each level of experience was examined using boxplots. There were no outliers in the unexpected engine failure flight or the expected engine failure for any of the three levels of expertise. The data were normally distributed for each flight type over each of the three levels of expertise (Shapiro-Wilk, $p > .05$). There was homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively.

A two way mixed ANOVA indicated that there was no significant interaction between expertise and flight type ($p = .827$). There was no significant main effect of expertise on mean pupil dilation ($p = .423$). However, there was a significant main effect of flight type on mean pupil dilation $F(2, 31) = 5.481$, $p = .026$, partial η2 = .146.

*Change in pupil dilation*

University students had the largest increase in pupil dilation (*M* = 2.20, *SD* = 0.68) from before the expected engine failure to during the expected engine failure, this was larger than the intermediate pilots (*M* = 1.28, *SD* = 0.51). Experts pilots (*M* = 1.55, *SD* = 1.00) had a smaller change in pupil dilation compared to the university students, and a larger increase in pupil dilation than the intermediate pilots (Figure 19). University students had the largest increase in pupil dilation (*M* = 2.66, *SD* = 1.09) from before the unexpected engine failure to during the unexpected engine failure, followed by the commercially licenced pilots (*M* = 1.97, *SD* = 0.55), and then the student and private licenced pilots (*M* = 1.69, *SD* = 0.96) who had the smallest increase pupil dilation (Figure 18).

The presence of outliers in the different change in pupil dilation data for each flight over each level of experience was examined using boxplots. There were no outliers in the expected engine failure flight for any of the three levels of expertise. In the unexpected engine failure the university students had one outlier, and the intermediate pilots had one extreme outlier and one regular outlier. The expert pilots had no outliers in the unexpected engine failure flights. The data were normally distributed for the expected engine failure over each of the three levels of expertise (Shapiro-Wilk, $p > .05$). The expert pilots also had normally distributed data for the unexpected engine failure (Shapiro-Wilk, $p > .05$). However the data were not normally distributed in the unexpected engine failure for the university students (Shapiro-Wilk, $p = .008$), or the intermediate pilots (Shapiro-Wilk, $p = .012$). Removal of all outliers led to a normal distribution in each flight over the difference expertise levels, and one further outlier that was assessed and kept in the analysis. There was homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively.

The two way mixed ANOVA indicated that there was no significant interaction between expertise and flight type on change in pupil dilation ($p$ =.905). There was a significant main effect of expertise on change in pupil dilation $F(2, 29) = 3.624$, $p = .039$, partial η2 = .200. Post hoc analysis with a Bonferroni adjustment revealed that university students had an increase in pupil dilation 0.76 pixels. 95% CI [0.03, 1.49] larger than intermediate pilots, a significant difference $p = 0.041$. There was no significant differences between expert pilots and university students or expert pilots and intermediate pilots. There was a marginally significant main effect of flight type on change in pupil dilation $F(2, 29) = 3.973$, $p = .056$, partial η2 = .120.

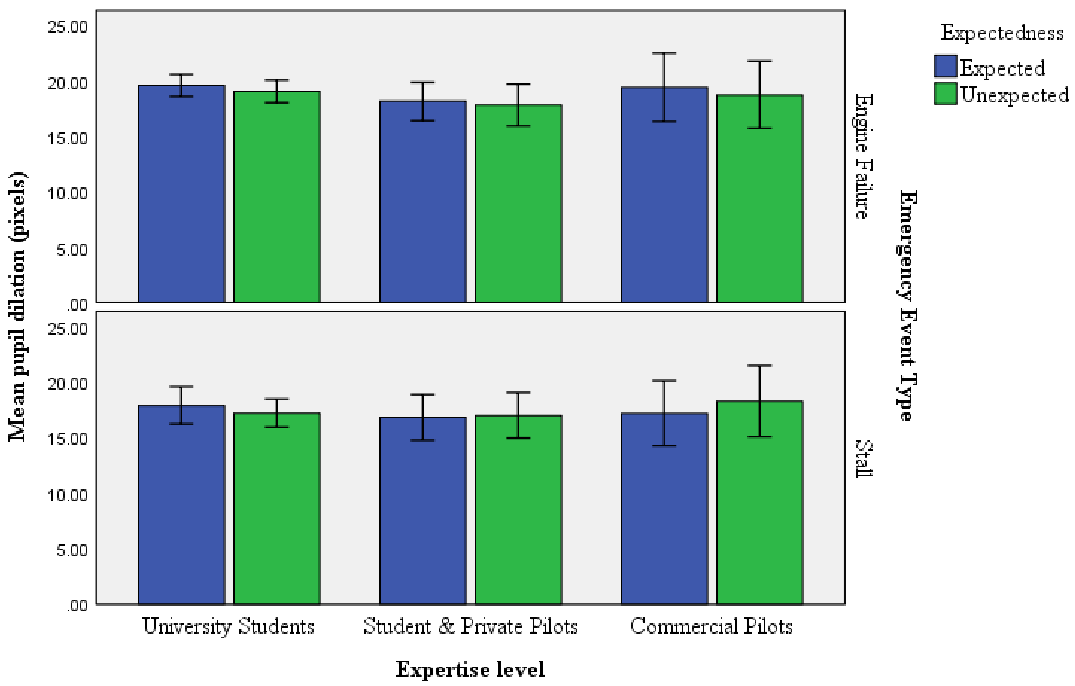



*Figure 18.* Mean pupil dilation (pixels) for university students, student and private (intermediate) pilots, and commercial (expert) pilots during the expected and unexpected engine failure and stall events, with error bars (+/- 2SE).

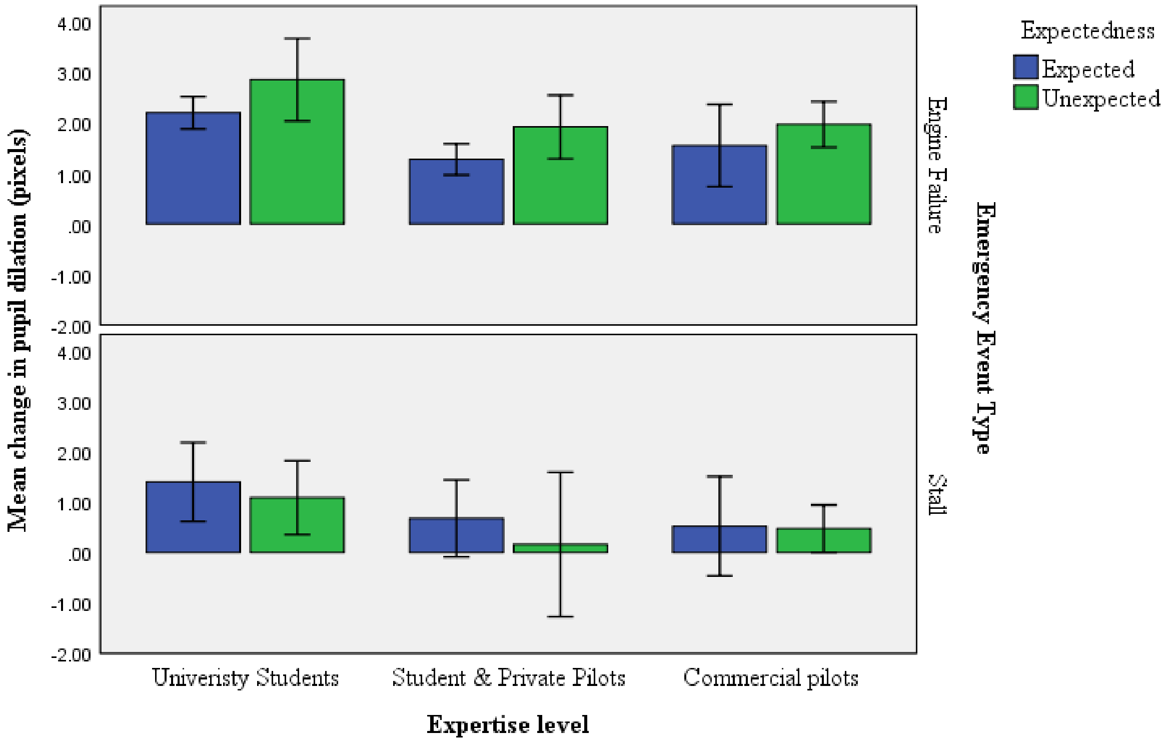



*Figure 19.* Mean change in pupil dilation (pixels) for university students, student and private (intermediate) pilots, and commercial (expert) pilots during the expected and unexpected engine failure and stall events compared to the baseline flight, with error bars (+/- 2SE).

***Stall***

*Mean pupil Dilation*

University students had the largest pupil dilation (*N*= 9, *M* = 17.86, *SD* = 0.84) in the expected stall, followed by the expert pilots (*N= 6*, *M* = 17.16, *SD* = 3.57), and then the intermediate pilots (*N= 7*, *M* = 16.80, *SD* = 2.73) who had the smallest pupil dilation (Figure 18). Expert pilots had the largest pupil dilation (*N*= 6, *M* = 18.24, *SD* = 3.91), in the unexpected stall followed by the university students (*N*= 9, *M* = 17.17, *SD* = 1.91), and then the intermediate pilots (*N= 7*, *M* = 16.97, *SD* = 2.73) who had the smallest pupil dilation (Figure 18).

The presence of outliers in the different change in pupil dilation data for each flight was examined over each level of experience using boxplots. There were no outliers in the unexpected stall flight for any of the three levels of expertise. In the expected stall the university students had two outliers. The intermediate pilots and the expert pilots had no outliers in the expected stall flight. The data were normally distributed for the expected and unexpected stall flights over each of the three levels of expertise (Shapiro-Wilk, $p > .05$). The two outliers were assessed and determined to be plausible and were therefore kept in the analysis. There was homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively.

A two way mixed ANOVA indicated that there was a significant interaction of flight type and expertise on mean pupil dilation $F(2, 19) = 3.668$, $p = .045$, partial η2 = .279. However, there was no significant main effect of flight type on mean pupil dilation ($p = .502$), and there was also no significant main effect of licence on mean pupil dilation ($p = .853$).

*Change in pupil dilation*

University students had the largest increase in pupil dilation (*M* = 1.40, *SD* = 1.17) from before the expected stall to during the expected stall, followed by intermediate pilots (*M* = 0.67, *SD* = 1.01), and then expert pilots (*M* = 0.52, *SD* = 1.20), who had the smallest increase pupil dilation (Figure 19). University students had the largest increase in pupil dilation (*M* = 1.08, *SD* = 1.10) from before the unexpected stall to during the unexpected stall, followed by the expert pilots (*M* = 0.47, *SD* = 0.57), and then the intermediate pilots (*M* = 0.15, *SD* = 1.89), who had the smallest increase in pupil dilation (Figure 19). A two-way mixed ANOVA indicated there was no significant interaction between flight type and expertise on change in pupil dilation ($p = .849$). There was no main effect of expertise on change in pupil dilation ($p = .206$), and there was no main effect of flight type on change in pupil dilation ($p = .367$).

**Information Processing**

Due to multiple testing occurring, an adjustment was made in so that only analyses where the ANOVA p-value as under 0.025 were considered significant. This value was chosen to prevent Type I and Type II errors.

***Engine Failure***

*Airspeed Indicator*

*Expected:* Intermediate pilots spent more time looking at the airspeed indicator ($N$ = 11, $M$ = 1.92%, $SD$ = 1.53) compared to expert pilots ($N$ = 6, $M$ = 1.71%, $SD$ = 1.11), and university students who spent the least percentage time viewing the airspeed indicator ($N$ = 18, $M$ = 0.33%, $SD$ = 0.36) (Figure 20). Due to the data not being normally distribution (positively skewed) Shapiro-Wilk test ($p > 0.05$), as well as outliers as assessed by boxplots, a square root transformation was completed. This resulted in a normal distribution of data and no outliers. Homogeneity of variance assumption was met as Levene's statistic was not significant; $p > .05$. A one-way ANOVA on the transformed data revealed a significant difference in the amount of percentage gaze time spent viewing the airspeed indicator over the three groups $F(2, 32) = 11.92$, $p < .001$. Post hoc Tukey tests indicated that the university students spent a significantly longer percentage of time viewing the airspeed indicator than the intermediate pilots ($p$ <.001), and the expert pilots ($p$ = .006). However, there were no significant differences between the expert pilots and the intermediate pilots.

*Unexpected:* Intermediate pilots spent more time looking at the airspeed indicator during the unexpected engine failure ($N$ = 11, $M$ = 0.90%, $SD$ = 0.68), compared to university students ($N$ = 18, $M$ = 0.60%, $SD$ = 0.63), and expert pilots ($N$ = 6, $M$ = 0.58%, $SD$ = 0.59), (Figure 20). A one-way ANOVA showed that these differences were not significant.

*Attitude Indicator*

*Expected:* University students spent a larger percentage of their time viewing the attitude indicator (*N* = 18, *M* = 0.70%, *SD* = 0.80) compared to intermediate pilots (*N* = 11, *M* = 0.28%, *SD* = 0.19), and expert pilots (*N* = 6, *M* = 0.28%, *SD* = 0.24) (Figure 20). A one-way ANOVA showed that these differences were not significant.

*Unexpected:* University students spent more time looking at the attitude indicator (*N* = 18, *M* = 0.62%, *SD* = 0.60) compared to intermediate pilots (*N* = 11, *M* = 0.17%, *SD* = 0.13), and expert pilots (*N* = 6, *M* = 0.10%, *SD* = 0.07) (Figure 20). Due to the data not being normally distribution (positively skewed) as well as outliers as assessed by boxplots and Shapiro-Wilk test ($p > 0.05$), a square root transformation was completed. This resulted in a normal distribution of data and no outliers. Homogeneity of variance assumption was not met as Levene's statistic $p = .007$. A one-way Welch ANOVA on the transformed data revealed a significant difference in the amount of time spent viewing the attitude indicator over the three groups Welch's $F(2, 20.251) = 6.373$, $p = .007$. Games-Howell post hoc tests showed that university students had significantly less viewing time than intermediate pilots ($p = .026$). University students also had significantly more viewing time than expert pilots, with a mean difference of ($p = .004$). However, there were no significant differences between the expert pilots and the intermediate pilots.

*Altimeter*

*Expected:* University students spend less time viewing the altimeter (*N* = 18, *M* = 0.21%, *SD* =0.28), compared to expert pilots (*N* = 6, *M* = 0.58%, *SD* = 0.80) and intermediate pilots (*N* = 11, *M* = 0.98%, *SD* = 0.98), who spent the most time viewing the altimeter (Figure 20). Due to the data not being normally distributed (positively skewed) as well as outliers as assessed by boxplots and Shapiro-Wilk test ($p > 0.05$), a square root transformation was completed. This resulted in a normal distribution of data and no outliers. Homogeneity of variance assumption was met as Levene's statistic $p = .163$. A one-way ANOVA on the transformed data revealed a significant difference in the amount of gaze time spent viewing the altimeter over the three expertise levels $F(2, 32) = 5.389$, $p = .010$. Post hoc Tukey tests indicated that university students spent significantly longer viewing the altimeter than the intermediate pilots ($p = .007$). However there were no significant differences between the expert pilots and the university students or the intermediate pilots.

*Unexpected:* Expert pilots spent less time looking at the altimeter ($N = 6$, $M = 0.15\%$, $SD = 0.20$), than the intermediate pilots ($N = 11$, $M = 0.26\%$, $SD = 0.17$), and the university students ($N = 18$, $M = 0.21\%$, $SD = 0.28$) (Figure 20). A one-way ANOVA showed that these differences were not significant.

*GPS*

*Expected:* Expert pilots spent more time looking at the GPS ($N = 6$, $M = 0.94\%$, $SD = 0.92$), than the intermediate pilots ($N = 11$, $M = 0.30\%$, $SD = 0.14$), and the university students ($N = 18$, $M = 0.17\%$, $SD = 0.21$) (Figure 20). Due to the data not being normally distributed (positively skewed) as well as outliers as assessed by boxplots and Shapiro-Wilk test ($p > 0.05$), a square root transformation was completed. This resulted in a normal distribution of data and no outliers. Homogeneity of variance assumption was met as Levene's statistic was not significant; $p > .05$. A one-way ANOVA on the transformed data revealed a significant difference in the amount of percentage gaze time spent viewing the GPS over the three groups $F(2, 32) = 10.719$, $p < .0005$. Post hoc Tukey tests indicated that expert pilots spent significantly longer viewing the GPS than the intermediate pilots ($p = .020$), and the university students ($p <.001$). There were no significant differences between the university students and the intermediate pilots.

*Unexpected:* Expert pilots ($N = 6$, $M = 0.47\%$, $SD = 0.40$) spent more time looking at the GPS than the university students ($N = 18$, $M = 0.21\%$, $SD = 0.28$), and the intermediate pilots ($N$ = 11, $M$ = 0.19%, $SD$ = 0.13), in the unexpected engine failure (Figure 20). Due to the data not being normally distributed (positively skewed) as well as outliers as assessed by boxplots and Shapiro-Wilk test ($p > 0.05$), a square root transformation was completed. This resulted in a normal distribution of data and no outliers. Homogeneity of variance assumption was met as Levene's statistic was not significant ($p > .05$). A one way ANOVA on the transformed data revealed a significant difference in the amount of percentage gaze time spent viewing the GPS over the three groups $F(2, 32) = 9.708$, $p = .001$. Post hoc Tukey tests indicate that the university students spent significantly longer viewing the GPS than the intermediate pilots ($p < .001$). However, there were no significant differences between the expert pilots and intermediate pilots, or the university students.

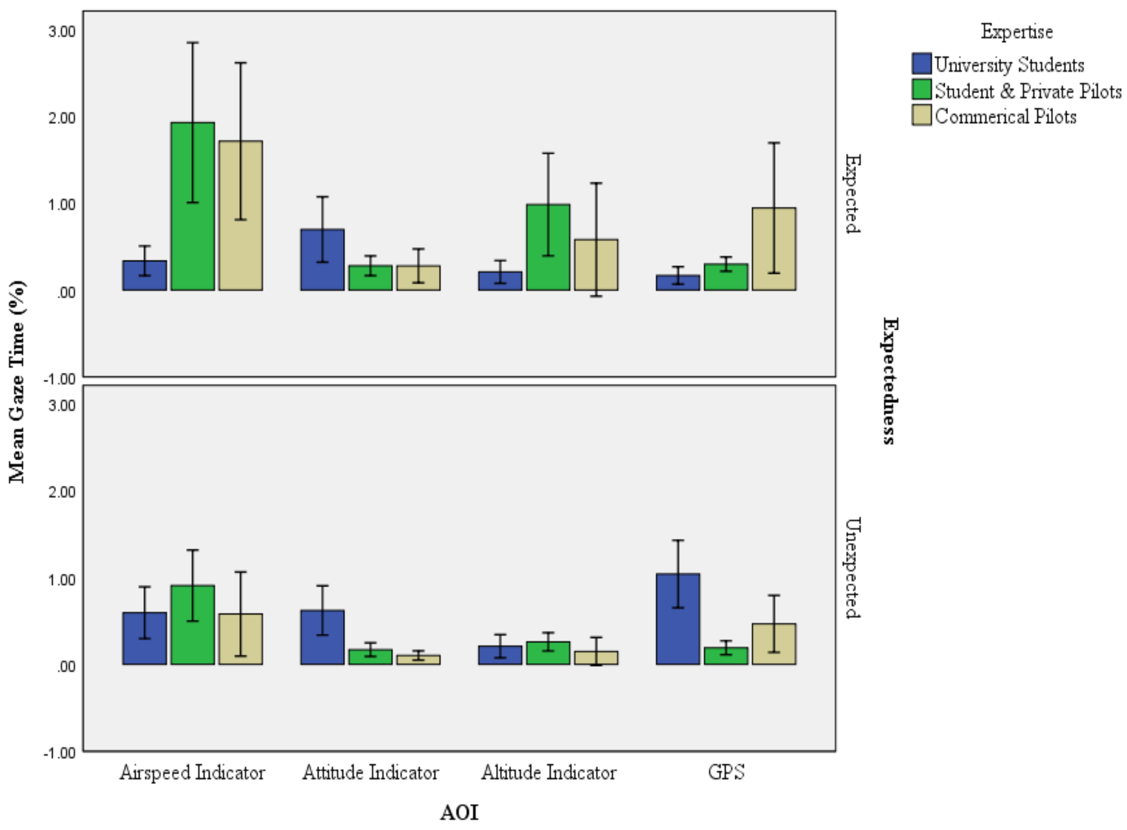



*Figure 20.* The mean percentage gaze time for the airspeed indicator, the attitude indicator, and altimeter, and the GPS AOI, for university students, student and private (intermediate) pilots, and commercial (expert) pilots for the expected and unexpected engine failures with standard error bars (+/-2SE).

*Inside*

*Expected:* University students spent less time looking at the PFD and the MFD on the LCD monitor (*N* = 18, *M* = 4.18%, *SD* = 3.11), than the intermediate pilots (*N* = 11, *M* = 6.59%, *SD* = 2.60), and the expert pilots (*N* = 6, *M* = 6.35%, *SD* = 1.21) (Figure 21). There was one outlier in the university student group, and the data were not normally distributed. When the outlier was removed, the data became normally distributed with no outliers. Homogeneity of variance assumption was met as Levene's statistic was not significant; $p = .072$. A one-way ANOVA on the transformed data revealed a significant difference in the amount of time spent viewing the PFD and MFD on the LCD monitor over the three groups $F(2, 31) = 9.204$, $p = .001$. Post hoc tests show that university students spent significantly less time viewing the flight display than intermediate pilots, with a mean difference of 3.02%, *SE* = 0.77, $p = .001$. University students also had significantly more viewing time than the expert pilots, with a mean difference of 2.78%, *SE* = 0.95, $p = .017$. Intermediate pilots spent more time looking at the flight display, compared to commercial pilots, however this difference was not significant ($p = .950$).

*Unexpected:* University students spent more time looking at the PFD and the MFD on the LCD monitor (*N* = 18, *M* = 4.87%, *SD* = 3.14), than the intermediate pilots (*N* = 11, *M* = 3.81%, *SD* = 1.56), and the expert pilots (*N* = 6, *M* = 3.57%, *SD* = 1.31) (Figure 21). A one way ANOVA showed that these differences were not significant.

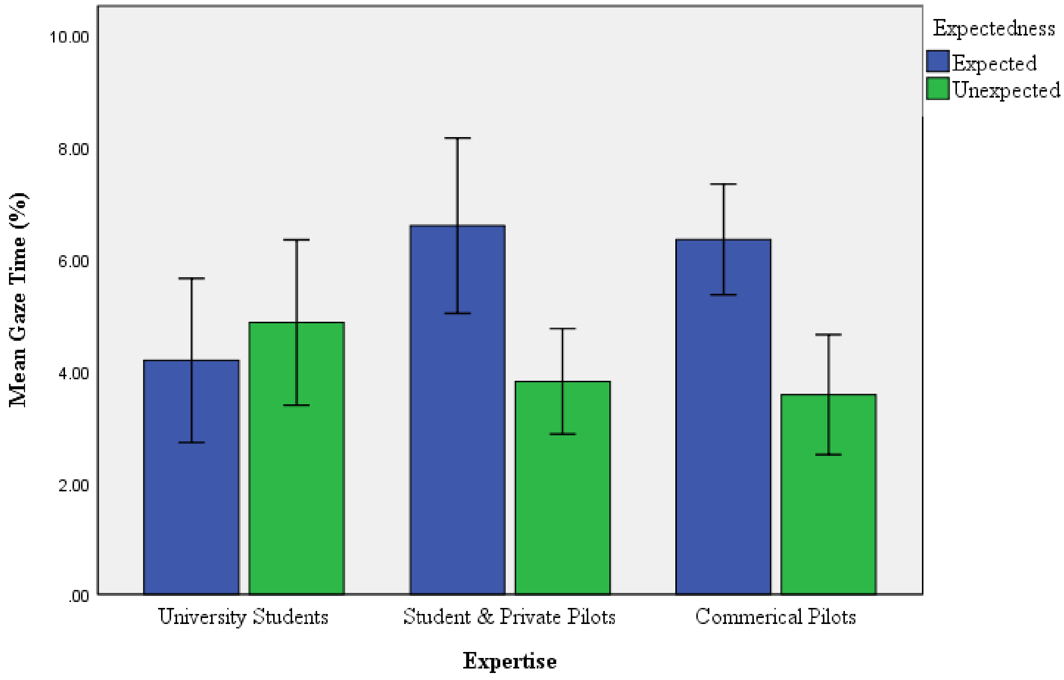



Figure 21. Mean percentage time university students, student and private (intermediate) pilots, and commercial (expert) pilots, spent viewing the overall PFD and MFD during the expected and unexpected engine failure event with standard error bars (+/-2 SE).

***Stall***

*Airspeed indicator*

*Expected:* Expert pilots spent less time looking at the airspeed indicator in the expected stall (*N* = 6, *M* = 0.11%, *SD* = 0.11), than both intermediate pilots (*N* = 7, *M* = 0.18%, *SD* = 0.09) and the university students (*N* = 9, *M* =0.21%, *SD* = 0.29) who spent the most time looking at the airspeed indicator (Figure 22). A one-way ANOVA showed that these differences were not significant.

*Unexpected:* Expert pilots spent more time looking at the airspeed indicator in the unexpected stall (*N* = 6, *M* = 1.24%, *SD* = 0.91), than the intermediate pilots (*N* = 7, *M* = 1.10%, *SD* = 1.22), and the university student (*N* = 9, *M* =0.15%, *SD* = 0.15), who spent the least time looking at the airspeed indicator (Figure 22). The intermediate pilot group had one outlier, but for all three groups the data was normally distributed. The outlier value of 3.61%, this was reduced down to the next largest value; 1.27%. Following this the data was normally distributed in each group, and there were no outliers as assessed by Shapiro-Wilk tests ($p <$ 0.05), and box plots respectively. The homogeneity of variance assumption was not met as Levene's statistic was significant ($p$ = .001). A one-way Welch ANOVA revealed a significant difference in the amount of time spent viewing the airspeed indicator over the three groups Welch's $F$(2, 7.713) = 7.528, $p$ = .015. However, Games-Howell post hoc tests show no significant differences between the groups.

*Attitude indicator*

*Expected:* University students spent the most time viewing the attitude indicator in the expected stall flight (*N* = 9, *M* = 0.13%, *SD* = 0.56), followed by expert pilots *(N* = 6, *M* = 0.09%, *SD* = 0.08), and then the intermediate pilots (*N* = 7, *M* = 0.04%, *SD* = 0.07) (Figure 22). A one way ANOVA showed that the differences were not significant.

*Unexpected:* University students spent the most time viewing the attitude indicator (*N = 9, M* = 0.34, *SD* = 0.47), followed by the expert pilot group *(N* = 6, *M* = 0.15%, *SD* = 0.09), and then the intermediate pilots (*N* = 7, *M* = 0.12%, *SD* = 0.09) (Figure 22). A one way ANOVA showed that these differences were not significant.

*Altimeter*

*Expected:* University students spent less time viewing the altimeter (*N* = 9, *M* = 0.01%, *SD* = 0.02), compared with the expert pilots *(N* = 6, *M* = 0.15%, *SD* = 0.22), and the intermediate pilots (*N* = 7, *M* = 0.15%, *SD* = 0.12) (Figure 21). A one way ANOVA showed that these differences were not significant.

*Unexpected:* Expert pilots spent a larger percentage of their time looking at the altimeter *(N* = 6, *M* = 0.46%, *SD* = 0.27), followed by the private and student pilots (*N* = 7, *M* = 0.13%, *SD* = 0.21), and then the university students (*N* = 9, *M* = 0.05%, *SD* = 0.06) (Figure 22). A one way ANOVA showed that these differences were not significant.

*GPS*

*Expected:* Intermediate pilots spent no time observing the GPS (*N* = 7, *M* = 0.00, *SD* = 0.00), compared to the university students (*N* = 9, *M* = 0.02%, *SD* = 0.03), and the expert pilots *(N* = 6, *M* = 0.03%, *SD* = 0.03), who both spent very little time viewing the GPS (Figure 22). A one way ANOVA showed that these differences were not significant.

*Unexpected:* In the unexpected stall university students spent the most time looking at the GPS (*N* = 9, *M* = 0.98, *SD* = 01.09), followed by the expert pilots (*N* = 6, *M* = 0.41, *SD* = 0.36), and the then private and student pilots (*N* = 7, *M* = 0.20, *SD* = 0.19) (Figure 22). A one way ANOVA showed that these differences were not significant.

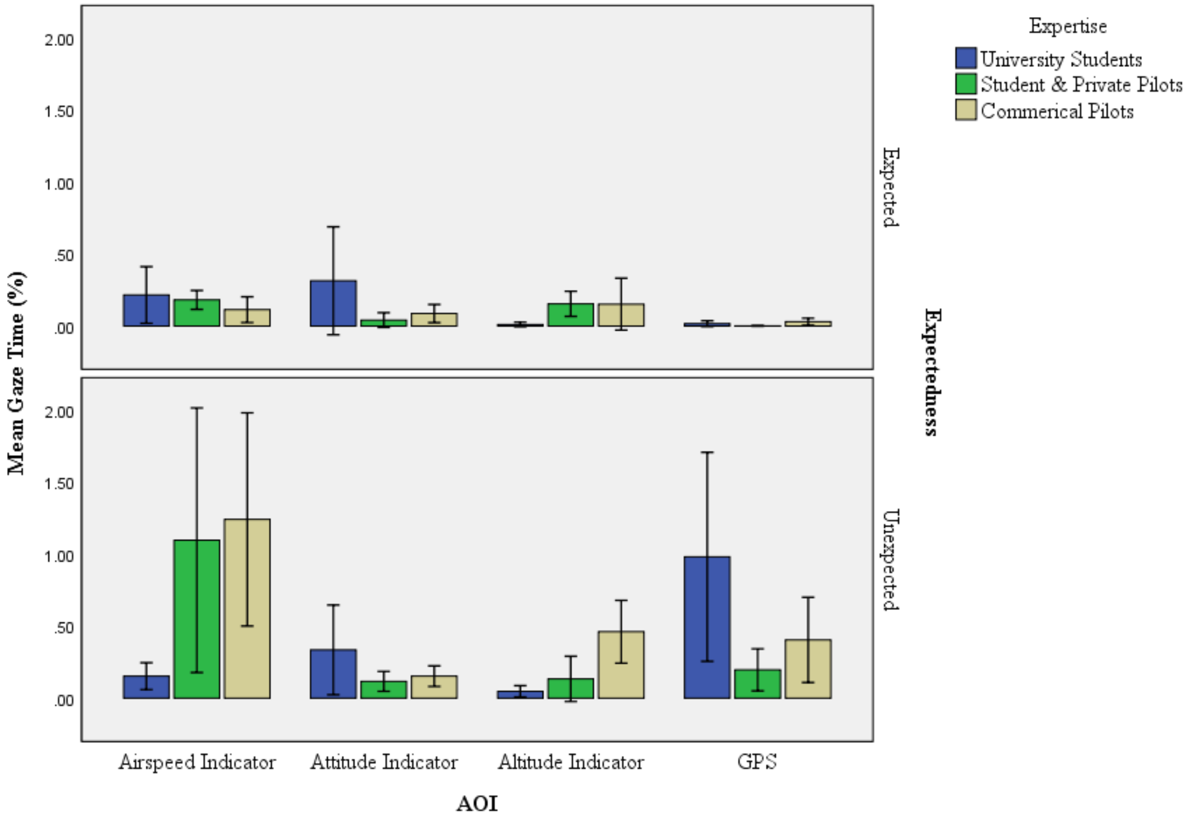



*Figure 22.* The mean percentage gaze time for the airspeed indicator, the attitude indicator, and altimeter, and the GPS AOI, for university students, student and private (intermediate) pilots, and commercial (expert) pilots for the expected and unexpected stall events, with standard error bars (+/-2SE).

*Inside*

*Expected:* University students spent more time looking at the PFD and the MFD on the LCD monitor (flight display)($N = 9$, $M = 1.04\%$, $SD = 0.86$), than the expert pilots ($N = 6$, $M = 0.96\%$, $SD = 0.34$) the intermediate pilots ($N = 7$, $M = 0.73\%$, $SD = 0.23$) (Figure 23). A one way ANOVA showed that these differences were not significant.

*Unexpected:* During the unexpected stall, expert pilots spent more time looking at the PFD and the MFD on the LCD monitor (flight display)($N = 6$, $M = 3.97\%$, $SD = 1.08$), followed by the intermediate pilots ($N = 7$, $M = 2.47\%$, $SD = 1.86$), and then the university students ($N = 9$, $M = 2.08\%$, $SD = 1.67$) (Figure 23). A one way ANOVA showed that these differences were not significant.

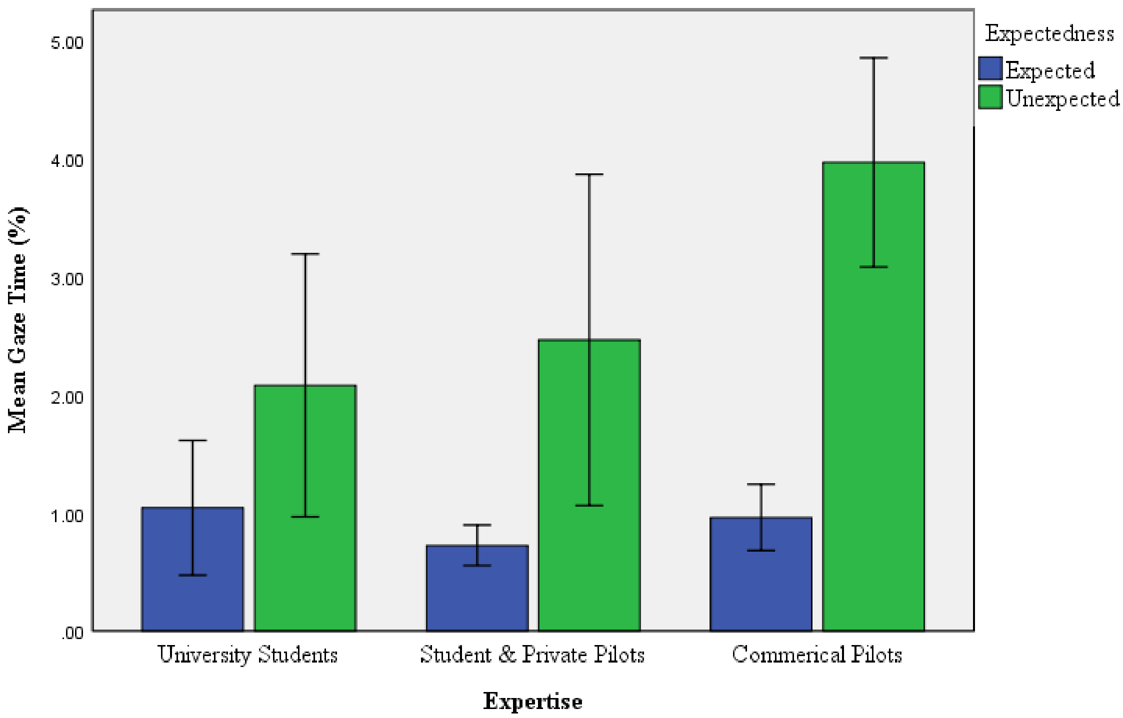



*Figure 23.* Mean percentage time university students, student and private (intermediate) pilots, and commercial (expert) pilots, spent viewing the overall PFD and MFD during the expected and unexpected aerodynamic stall events, with standard error bars (+/-2 SE).

**Expertise and Performance**

***Engine Failure***

*Crash or Safe landing*

Four out of nine (44.40%) expert pilots, six out of 13 (46.20%) of intermediate pilots, and 14 out of 21 university student (66.70%) crashed after the unexpected engine failure, or attempted to land in Lake Wakatipu. A chi squared analysis showed no significant difference in the distribution of pilots that crashed and landed safely in the unexpected engine failure flight over licence type. However, over 50% of both pilot groups were able to land safely where only around one third of university students were able to land safely.

*Retard throttle*

Half (50%) of expert pilots (3 of 6) did not pull back on the throttle in both the expected and unexpected engine failure. Where 3 out of 13 intermediate pilots (23.1%) did not pull back on the throttle in both the engine failure flights. Expert pilots that did remember to pull back on the throttle did so faster in the expected ($N = 3$, $M = 1.33$ seconds, $SD = 0.58$) compared to the unexpected engine failure ($N = 3$, $M = 2.00$ seconds, $SD = 1.00$. Intermediate pilots that did remember to pull back on the throttle did so faster in the expected ($N = 9$, $M = 3.11$seconds, $SD = 2.67$) compared to the unexpected engine failure ($N = 9$, $M = 3.89$seconds, $SD = 2.20$). Intermediate pilots took longer than commercial pilots to pull back on the throttle in both flights. Due to the very small difference in sample size these differences are not comparable.

***Stall***

*Response*

One of eight (12.50%) university students, two of seven (28.57%) intermediate pilots, and two of six (33.33%) expert pilots incorrectly pulled back on the throttle following stall.

Five of eight (62.50%) university students, four of seven (57.14%) intermediate pilots, and one of six (16.67%) of expert pilots did not pitch the plane down following the stall horn.

***Recovery time***

*Expected:* The time to recover from the expected practice stall was calculated from the time that the participant retarded the throttle, and the airspeed was reduced to 60 knots, then after the stall occurred till the plane returned to level flight. University students spent less time recovering from the expected stall ($N = 7$, $M = 13.86$seconds, $SD = 3.48$), compared to intermediate pilots ($N = 7$, $M = 22.42$seconds, $SD = 16.19$), and expert pilots ($N = 6$, $M = 19.71$seconds, $SD = 3.17$).

*Unexpected:* The time to recover from the unexpected stall was calculated from the time that the plane reached 500ft till the time that the participant successfully passed the wind shear and reached 550ft. University students spent less time recovering from the unexpected stall ($N = 7$, $M = 19.71$seconds, $SD = 7.61$), compared to intermediate pilots ($N = 7$, $M = 27.48$seconds, $SD = 8.07$), and expert pilots ($N = 6$, $M = 47.20$seconds, $SD = 24.47$).

***Altitude lost***

*Expected:* University students lost a large amount of altitude in the expected stall ($N = 7$, $M = 918.57$feet, $SD = 537.94$), compared to intermediate pilots ($N = 7$, $M = 192.14$feet, $SD = 56.71$), and expert pilots ($N = 6$, $M = 241.67$feet, $SD = 86.35$).

*Unexpected:* In the unexpected stall the expert pilots lost the most altitude ($N = 6$, $M = 81.67$feet, $SD = 89.54$), followed by the university students stall ($N = 7$, $M = 63.57$feet, $SD = 113.53$), and then the intermediate pilots ($N = 7$, $M = 40.71$feet, $SD = 37.46$).

## Discussion

Due to very small sample sizes this third comparative study is mostly exploratory in nature. The small sample sizes reduce the probability of finding any significant differences. However, this exploratory study may provide some insights which could lead to further research initiatives. It would seem to be a wasted opportunity if the comparisons between the three levels of expertise were not made. However, this third investigation and the tentative conclusions made would clearly have benefited by having more pilot participants.

### Physiological measures

It was hypothesized that during the expected and unexpected engine failure and stall flight events, expert pilots would have the lowest mean heart rate when compared to intermediate pilots and novices. It was further hypothesized that novices would have larger mean heart rates when compared to intermediate pilots.

There was a significant interaction between flight type and experience level on mean heart rate engine failure flights as well as baseline. Post hoc analyses indicated that contrary to the hypothesis there was no statistically significant effect of expertise on heart rate during the expected engine failure or the baseline flight. There was a significant difference in heart rate over the three levels of expertise for the unexpected engine failure. However, contrary to the hypothesis the intermediate pilots had a significantly higher heart rate compared to the university students (novices) in the unexpected engine failure. There were no further significant differences between expertise levels in the unexpected engine failure.

Consistent with the hypothesis there was a significant overall effect of expertise on change in heart rate in the engine failure flights compared to baseline. Contrary to the hypothesis university students had significantly smaller mean change in heart rate than both the intermediate and expert pilots.

There was no significant interaction between flight type and expertise on mean heart rate during the expected and unexpected stall and the baseline flight. There was only no significant main effects of flight type or expertise. There was no significant interaction between flight type and expertise on change in mean heart rate compared to baseline during the stall flights. There was also no significant main effects of expertise or flight type on change in heart rate in the stall flights.

Although there was a lack of significant differences in mean heart rate between the levels of expertise, over the different event flight types, there are speculations that can be made on some of the data trends as shown in Figure 17. Intermediate pilots had the highest heart rate in all experimental flights. While university students had the lowest average heart rate in all flights apart from the expected stall. Expert pilots had a lower heart rate than intermediate pilots in all flights and a slightly lower heart rate than the university students in the expected stall flight. Furthermore, both intermediate pilots and university students had a decrease in heart rate in the unexpected stall event compared to baseline, while expert pilots had an increase. Expert pilots show an autonomic arousal trend consistent to what was expected, a low heart rate in the baseline flight and expected versions of the two events and a higher heart rate in the unexpected engine failure and stall (Figure 17).

It was hypothesized that during the engine failure and stall flight events, experts would have the smallest pupil dilation and change in pupil dilation compared to intermediate pilots and novices, with novices having the largest. Inconsistent with the hypothesis, there no significant interaction between flight type and expertise on mean pupil dilation for the engine failure events. There was also no significant main effect of expertise on mean pupil dilation in the engine failure flights. There was however a significant effect of flight type on mean pupil dilation for the engine failure flights.

Inconsistent with the hypothesis, there no significant interaction between flight type and expertise on mean change in pupil dilation for the engine failure events. There was also no significant main effect of flight type on change in mean pupil dilation in the engine failure flights. There was however a significant effect of expertise on change in mean pupil dilation for the engine failure flights. University students had a significantly larger change in pupil dilation during the engine failure compared to intermediate pilots.

There was a significant interaction between flight type and expertise on mean pupil dilation in the stall flights. However there was no significant main effect of expertise or flight type on mean pupil dilation in the stall flights. There was no significant interaction or main effects of expertise or flight type on change in pupil dilation in the stall flights.

There was a lack of significant differences in pupil dilation and change in pupil dilation over the three different levels of expertise. However when looking at the non-significant trends, concordant with the hypothesis university students had the largest pupil dilation and increase in pupil dilation in both the expected and unexpected engine failures and the expected stall flight. Contrary to the hypothesis expert pilots had a larger average pupil dilation and increase in pupil dilation compared to the intermediate pilots in both the expected and unexpected engine failures. Furthermore, contrary to the hypothesis, expert pilots had the largest pupil dilation in the unexpected stall, followed by the university students an then the intermediate pilots.

Consistent with heart rate data both intermediate and expert pilots show an increase in pupil dilation for both engine failure types, with the unexpected engine failure leading to a larger increase in pupil dilation. The very slight increase in pupil dilation for the intermediate pilots in the unexpected stall is consistent with the heart rate findings. Furthermore expert pilots also show similar increases in pupil dilation in both expected and unexpected stalls, which is also consistent with the heart rate findings. University students had larger changes in pupil dilation when compared to both pilot types, which is the opposite of the heart rate data. Therefore both the heart rate and pupil dilation have a consistent pattern of autonomic arousal for the pilots. However, university students shown larger increases in pupil dilation and very low differences in heart rates in respect to the two pilot groups. This indicates that there may be a variable or attribute affecting pilots heart rates as well as pupil dilation that is not affecting the university students.

The following interpretation of these non-significant data is tentative at best. The differences in heart rate may be due to prior experience with the abnormal events and in a flight simulator. Some intermediate pilots had relatively little previous flying experience (i.e. four pilots had less than 30 hours). It is possible that the expert pilots' higher level of experience allowed them to feel more comfortable in a flight simulator compared to the intermediate pilots. This is supported by the trend that intermediate pilots had a higher mean heart rate in the baseline flight than the expert pilots. Furthermore, completing a power-off stall practice without a prior refresher may have been stressful as the event would not have been as well-practiced compared to the pilots with more flight hours. The expectation of remembering and implementing the correct procedures consequently may have been demanding leading to the higher mean heart rate for the intermediate pilots during the expected stall compared to the expert pilots. Intermediate pilots had a heart rate in the unexpected stall similar to during their baseline flight. This may suggest that they did not recognize that an event was occurring. In contrast expert pilots showed an elevated heart rate compared to baseline. This may suggest that they did recognize the stall event.

Increases in heart rate are well accepted to be due to the activation of the autonomic nervous system (LeDoux, 2003). An increase in heart rate is indicative of startle. (Chou et al., 2014; Holand et al., 1999), stress (Lahtinen et al., 2007), and cognitive workload (Fallahi et al., 2016; Grassmann et al., 2017). Additionally, pupil dilation has been shown to increase during the startle reaction (Bradley et al., 2005; Rivera et al., 2014), with autonomic arousal, and with increasing cognitive workload (Bradley et al., 2008; Einhäuser et al., 2008; Hyönä et al., 1995; Kahneman, 1973; Marinescu et al., 2018; Marshall, 2002). The non-significant data suggest that the intermediate pilots found all flights apart from the unexpected stall and baseline flight stressful or cognitively demanding, with the expected stall having the largest degree of physiological arousal. Expert pilots elicited high physiological arousal suggesting high cognitive workload or stress in both the engine failures and the unexpected stall, with the unexpected engine failure leading to the most autonomic arousal. The university students showed noticeable increases in heart rate for only the expected stall.

Therefore compared to the other two groups the expert pilots appear not to have had a physiological response to the expected stall, suggesting that their experience has led to low workload or a lack of stress during this event. The arousal in the expected engine failure could be due to the fact that unlike normal training the participants were required to land. Lahtinen et al. (2007) showed that combat pilots had an increase in heart rate of around 5bpm above baseline during the routine parts of their flight e.g. take-off and landing. Therefore landing is likely to increase workload, which may be reflected by an elevated heart rate.

Computer gaming research has investigated and characterised immersion (AKA presence) (Alexander, Brunyé, Sidman, & Weil, 2005; Brown & Cairns, 2004; Cheng & Cairns, 2005; Witmer & Singer, 1998). In terms of gaming, immersion describes the level of involvement, absorption, or engrossment a person feels when playing a game (Brown & Cairns, 2004). There is a plethora of research concerning immersion which is beyond the scope of this thesis (see Cairns, Cox, and Nordin (2014) for a recent review). Brown and Cairns (2004) conducted a qualitative study investigating game immersion. They characterized three different levels of immersion which are now widely used; engagement, engrossment, and total immersion (ascending order). An individual needs to pass the first level to enter the second and so forth. The lowest level of immersion, engagement requires time, effort, and attention where the individual learns the game and its controls. The second level is engrossment, where an individual is familiar with the game, their emotions are directly affected by the game. and they have less awareness of their surroundings (Brown & Cairns, 2004; Cheng & Cairns, 2005). The third level is total immersion whereby individuals are further involved and completely focused in the game's reality, their awareness of the outside world is minimal (Brown & Cairns, 2004; Cheng & Cairns, 2005). Research has indicated that higher levels of perceived control over the game leads to higher subjective immersion (Sadowski & Stanney, 2002; Witmer & Singer, 1998). Extrapolating these findings to the flight simulator set-up used in the current experiment, it would suggest that piloting experience or lack of piloting experience would affect immersion.

University students are likely to only be operating in the engagement stage of immersion during the experiment. Before undergoing the experiment, the students had simply been introduced to the simulator and had learnt the bare minimum skills required to pilot an aircraft without major issues. Therefore it is unlikely students would have felt high levels of control during the flights, which in turn may have meant low levels of immersion (Sadowski & Stanney, 2002; Witmer & Singer, 1998). MSFX is a computer game, and according to the three levels of immersion the participants were likely only at the engagement stage of immersion during the flight tasks, where their emotions would have been minimally affected. Low immersion in the MFSX would lead to a low probability of a startle reaction, or emotionally driven autonomic arousal, and therefore could explain the small, non-significant changes in heart rate.

As both heart rate and pupil dilation are functions of the autonomic nervous system, it is thought that their patterns of arousal in participants would align. However, this was not the case for university students. A very tentative theory behind this is that elevated heart rate was not increased due to lack of immersion and then lack of emotional response by university students in the flight simulator. Brosschot and Thayer (2003) found that heart rate response is increased following negative emotions. It is possible that the university students have a high workload following the commencement of both the expected and unexpected stalls and engine failure. This may have been reflected by their increases in pupil dilation. However they are not emotionally responsive to the flight simulator, therefore their heart rate response may be shorter than the pilot groups.

It has been suggested that lack of variable training for emergency events, could lead to an underdevelopment of true expertise, where pilots would not be able to generalise their skills to an unexpected flight event (Casner et al., 2012). In the present study it was found that pilots with a higher number of flight hours (>100), had a lower heart rate than intermediate pilots in all flights. This suggests that with more experience, pilots may develop coping skills allowing them to be less reactive in emergency events. Furthermore, while both intermediate pilots and university students did not show a large response to the unexpected stall, expert pilots' data suggests an increase in pupil dilation and heart rate compared to baseline. This suggests that the expert pilots may be better at recognizing when an event is occurring i.e. have better awareness of the plane's condition. This is a promising finding as it supports the current training paradigm as well as suggesting that expertise does promote situational awareness in pilots and their ability to help pilots maintain focus during a real emergency.

**Information processing**

It was hypothesized that expert pilots would spend less over all time looking inside the cockpit at the MFD and PFD, than the intermediate pilots during the unexpected engine failure and stall. This is because expert pilots should be faster at diagnosing flight situations and therefore would be quicker to implement an effective response and information search protocol. Additionally, Wiggins and O'Hare (1995) found intermediate pilots appear to try and gather as much information during pre-flight decision making. Furthermore, it was hypothesized that university students would spend the most time looking at the flight display in the unexpected engine failure as they would need to spend more time diagnosing the situation. University students also were not taught to focus on the external environment during engine failures. Compared to both pilots groups, it was hypothesized that novices would spend the least amount time viewing the flight display (MFD and PFD) during the unexpected stall, as it was likely that they would not identify that there is a flight event.

Contrary to the hypotheses there were no significant differences in the amount of time each different group spent viewing the flight displays in the unexpected engine failure. However when examining the trends, consistent with the hypothesis, university students spent more time viewing the flight displays (MFD and PFD) in the unexpected engine failure than the two groups of pilots. Contrary to the hypothesis in the unexpected engine failure both groups of pilots spent a similar amount of time viewing the flight displays. This is possibly because an engine failure is quite a salient abnormal event, which is easily recognizable by any pilot.

The investigation into time spent by participants looking at the different displays is exploratory research. No specific hypotheses were made regarding this section of data. It is also recognized that multiple comparisons are occurring. Therefore data not significant at the one percent ($p$ <.025), were not considered as significant. During the unexpected engine failure the university students spent significantly more time viewing the attitude indicator compared to the two pilot types. University students also spent significantly more time viewing the GPS compared to the intermediate pilots in the unexpected engine failure. Therefore, the data indicate that there are detectable differences in the information processing strategies used by novices compared to pilots in abnormal events.

There were no significant differences in time spent looking at the different displays in the unexpected engine failure between the two pilot groups. Thus the following differences reported were non-significant and therefore could have occurred by chance. Intermediate pilots spent a longer percentage time viewing the airspeed indicator, the attitude indicator, and the altimeter in the unexpected engine failure compared to the expert pilots. The expert pilots spent longer viewing the GPS in the unexpected engine failure compared to the intermediate pilots. In the expected engine failure, expert pilots spent significantly longer viewing the GPS. The non-significant findings show that intermediate pilots spent slightly longer viewing the airspeed indicator and the altimeter, and a similar time viewing the attitude indicator and the entire flight display (MFD and PFD) compared to the expert pilots in the expected engine failure.

In the expected and unexpected engine failures intermediate pilots spent slightly more time viewing the airspeed indicator and the altimeter. This could possibly indicate less efficiency at extracting the required information. Expert pilots spent longer looking at the GPS in both flights; this may be because they were trying to maintain their situational awareness by compensating for the lack of ability to look out the side windows. Or because the expert pilots probably have more exposure to cross-country flights they have developed a higher reliance on the GPS compared to pilots that are still in flight school. Either possibilities are merely thoughts and suggestions.

During the unexpected stall there was a significant effect of expertise on time spent viewing the airspeed indicator, however post hoc tests showed no significant differences. Expert pilots viewed the airspeed indicator more than the intermediate pilots and the university students. University students spent the least amount of time viewing the airspeed indicator. The rest of the instruments as well as the entire flight display (MFD and PFD) showed no significant differences in view time over the three groups of expertise.

The following differences were not significant, however they are mentioned because there were noticeable differences between the groups, suggesting that with more power, real differences could be found. When examining the trends, consistent with the hypothesis university students spent less time viewing the flight display (MFD and PFD) in the unexpected stall than the two pilot groups. However, contrary to the hypothesis expert pilots spent more time looking inside at the MFD and PFD during the unexpected stall, than the intermediate pilots. In the unexpected stall expert pilots spent longer looking at the altimeter, and the GPS than the intermediate pilots. In the unexpected stall university students spent more time viewing the attitude indicator, and the GPS than both classes of pilots. In the unexpected stall university students spent less time viewing the airspeed indicator, and the altimeter than both classes of pilots. In comparison, during the expected stall intermediate pilots and expert pilots spent a similar amount of time viewing the airspeed indicator, the attitude indicator, the altimeter, and the GPS. Expert pilots spent slightly more time overall viewing the flight displays.

In the expected stall both pilot groups had a similar pattern of information search. In the unexpected stall expert pilots spent more time viewing stall relevant instruments, e.g. the altimeter and the overall flight display. Therefore in the unexpected stall there were discrepancies that support the theory that expert pilots had more awareness that there was an abnormal event occurring.

**Performance**

It is hypothesized that expert pilots would perform the best compared to intermediate pilots and university students, and intermediate pilots would perform better than university students. Concordant with the hypothesis a smaller percentage of expert pilots crashed in the unexpected engine failure, compared to intermediate pilots and university student. Additionally university students had the highest crash rate. These differences were not compared due to small sample sizes. Contrary to the hypothesis, a higher percentage of expert pilots did not pull back on the throttle in the expected and unexpected engine failure compared to the intermediate pilots. However those expert pilots that did pull back the throttle had a lower mean response time to than the intermediate pilots in both engine failure flights.

One of eight (12.50%) university students, two of seven (28.57%) intermediate pilots, and two of six (33.33%) expert pilots incorrectly pulled back on the throttle following stall. Five of eight (62.50%) university students, four of seven (57.14%) intermediate pilots, and one of six (16.67%) expert pilots did not pitch the plane down following the stall horn in the unexpected stall. In the expected stall university students took the fastest time to return to stable flight after implementing a stall but lost the most altitude. Commercial pilots were the fastest to recover in the expected stall, but lost more altitude than the intermediate pilots. In the unexpected stall university students had the lowest recovery time, followed by the intermediate pilots, and then the expert pilots. In the unexpected stall the commercial pilots lost the most altitude, followed by the university students, and then the intermediate pilots.

Performance over the three expertise groups was variable. The performance results support the idea that expert pilots recognized that there was an abnormal event occurring in the unexpected stall, as a higher percentage pulled back on the throttle and pitched the nose downwards. Although pulling back on the throttle was the incorrect response, it is still a salient indicator that they were aware that there was something to respond to. A very large percentage of intermediate pilots did not pitch the nose down following the stall horn, which supports the hypothesis that intermediate pilots would perform more poorly than the expert pilots. As well as this intermediate pilots lost the least altitude in the unexpected stall, this is due to them not pitching the nose downwards or adjusting their flying following the stall horn. Although the results were optimal in this case, this was only due to the computer programming of the stall. If it had occurred in a real life scenario and the pilots did not alter their trajectory, or respond to the stall horn an accident could have occurred.

The definition of intermediate pilots and expert pilots in the current study may not be accurate in investigating expertise. O'Hare et al. (1994) and Wiggins and O'Hare (1995) defined novice pilots as having under 100 hours, intermediate pilots as having between 100 and 1000 hours, and expert pilots as those having over 1000 hours. Furthermore they found that among aircraft accidents in New Zealand more intermediate level pilots were involved in serious aviation accidents. However, in this present study intermediate pilot were defined as having between 1 and 250 hours, whereas expert pilots were pilots that had over 250 hours. Only three expert pilots had over 1000 hours. Therefore the differences in expertise may not have been prominent as the sample in the present study was weighted towards less experienced pilots.

Expertise is generally and subjectively defined by factors such as reputation or amount of time spent within a domain (Boot & Ericsson, 2013). However these definitions may fail in identifying individuals with genuine exceptional performance (Boot & Ericsson, 2013; Loveday, Wiggins, Festa, Schell, & Twigg, 2013). A more valid measure of an individual's skill can come from observing his or her objective performance (Boot & Ericsson, 2013). The expert performance approach focuses on identifying tasks that reliably discriminate levels of performance within a domain (Boot & Ericsson, 2013). O'Hare (2003) has reported that the quality of pilot decision making depends on many factors, a critical factor being recent and/or specific flying experience rather than total number of hours. Flying experience is thought to improve a pilot's ability to understand and respond to problems (O'Hare, 2003). It is possible other performance factors, behavioural tendencies, or pilot characteristics may have been better suited to define expertise e.g. recent flight hours.

**Conclusion**

In conclusion this third comparative study investigating the effects of expertise has some tentative and preliminary findings that with an increase in expertise, comes a decrease in autonomic arousal associated with flying and encountering abnormal events. Both intermediate pilots and expert pilots had spent a similar amount of time viewing the flight displays and different instruments, however intermediate pilots data may tentatively suggest less efficiency processing information compared to the expert pilots. Furthermore, the data suggests that pilots with more experience are better able to recognize an important aircraft upset and respond where less experienced pilots may not respond to flight events they do not recognize. Overall the comparisons made in this third study were mostly not significant. However, the patterns in the data may pique other researcher's interests and therefore are worth discussion.

# Chapter 5: General Discussion and Conclusions

## Research Questions:

### *Do simulated unexpected emergency flight events cause physiological startle?*

The findings in the current study support the idea that unexpected events in a flight simulator can lead to physiological arousal that is consistent with startle (Bradley et al., 2005; Chou et al., 2014; Deuter et al., 2012; Holand et al., 1999; Lahtinen et al., 2007; Rivera et al., 2014). However, physiological responses similar to startle were only found in response to an unexpected engine failure, but not to an unexpected stall. An unexpected stall due to sudden changes in wind direction, is not an emergency unless it is not recovered from. Therefore the lack of startle, is in line with Landman et al.'s (2017a) conceptual model of startle and surprise, as startle only occurs after perception of an abnormality when intensity or threat reaches a certain threshold (Figure 1; as reproduced in Study 2 Discussion). The unexpected stall in the present study recovered when the pilots flew below 500 feet or above 550feet. This may have impacted on the intensity or the threat of the unexpected stall impeding the probability of the stall escalating to startle.

A recent theory proposed and investigated by a number of researchers is that startle following the commencement of an unexpected abnormal flight event is a major factor in aviation accidents (Landman et al., 2017a, 2017b; Martin et al., 2015; Martin et al., 2010; Martin et al., 2016; Schroeder et al., 2014). Similar to other stress types, research shows that stress from startle can degrade information processing (Eysenck et al., 2005; Thackray & Touchstone, 1983), reduce working memory capacity (Bradley et al., 2005) and produce decisional errors (Driskell & Salas, 2013; May & Rice, 1971; Staal, 2004). It is thought that cognitive impairment that follows startle leads to poor performance in what should be well-practiced flight events (Landman et al., 2017a, 2017b; Martin et al., 2015; Martin et al., 2010;

Martin et al., 2016; Schroeder et al., 2014). This is also supported by findings in aviation accident reports and reviews of accident causations (NTSB, 2010a; Rivera et al., 2014). A limitation of the previous empirical research investigating this phenomena is that it commonly used loud noises or distraction to induce startle (Landman et al., 2017b; Martin et al., 2016). Distraction is also known to cause impairments in flight performance (Barnes & Monan, 1990), and is therefore a confounding factor. The present study provides physiological evidence that startle and surprise can occur due to unexpectedness i.e. without a loud noise or distraction. Therefore, the present study supports the previous research, and the theory that startle due to the unexpectedness of the emergency event could be a factor in performance degradation in unexpected abnormal flight events.

This is a promising finding as it also indicates that startle can be produced in a simulated environments. The present study used a fixed-base simulator, which would be considered low-fidelity in terms of modern simulator technology. Pilots may be more immersed in higher fidelity simulators, therefore a more pronounced startle effect may be found (see Alexander et al. (2005) for a review on fidelity, immersion, and presence in terms of training applications). Previous research has shown that startle produced by a loud noise can disrupt a participant's initial reaction in a flight-related task (Thackray, 1965). However, with further exposure subjects' performance in the flight related task improved following the startling stimulus (Thackray, 1965). It appears imperative to prepare pilots for unexpected, unusual and distracting events to enhance their ability to recover from them. Concordant with this, reviews of experiments, surveys, and accident and incident reports by Green (1895) indicated that pilots that dealt well with acute stress attributed their performance to simulation training (Green, 1985). Therefore, simulator exposure to variable unexpected events could be used in training to try and extinguish fear potentiated startle in response to real-life unexpected events.

Future research could investigate simulated abnormal events and characterise which events have a high probability of leading to arousal indicative of the startle reaction in pilots. This would provide a good substrate for training and future research into startle following unexpected flight events. A great place to start with that research would be Martin et al.'s (2015) formulated list of potential situations that would lead to startle. However, it would also be beneficial for future research to find abnormal events that lead to startle by simply occurring unexpectedly as opposed to having too many confounding factors e.g. distraction or erroneous instructions.

***What are the differences in information processing when comparing responses to an expected emergency event and an unexpected emergency event?***

The present research indicates that there are significant differences in fixation patterns in expected compared to unexpected events. In the unexpected engine failure pilots spent significantly less time viewing all of the flight instruments and the combined MFD and PFD compared to the expected engine failure. In the unexpected stall pilots spent a significantly longer amount of time viewing the flight instruments and the combined MFD and PFD compared to the expected stall. Therefore, this research provides evidence that the pattern of information processing is different when comparing expected verses unexpected versions of the same abnormal flight event. However, the nature of the changes in information gathering and processing are event specific.

An engine failure is a highly salient event, therefore pilot's information processing change may be similar to attentional tunnelling. The pilots had fewer fixations on the flight instruments (peripheral cue utilisation) and more focus on the external environment (the threat) in the unexpected engine failure. This indicates a change in informational processing analogous to attentional tunnelling. Attentional tunnelling involves a reduction in peripheral cue utilisation, and a narrowing of the attentional field towards the threat (Baddeley, 1972;

Combs & Taylor, 1952; Easterbrook, 1959; Staal, 2004). Furthermore, attentional tunnelling has been found to produce decisional errors (Driskell & Salas, 2013; May & Rice, 1971; Staal, 2004). The present findings are consistent with statements from the FAA indicating that in high stress situations such as emergency events, pilot information scan can be severely reduced (FAA, 1988). Additionally the current study supports previous research which indicates that startle can lead to disruptions in information processing, as well as extending it to a complex operational situation. (Baddeley, 1972; Combs & Taylor, 1952; Easterbrook, 1959; Staal, 2004).

It is postulated that the pilots did not recognize the stall or were confused about its origin, leading to a longer appraisal period. The stall led to information processing in line with the hypothesis, which stated that the pilots would need to spend more time diagnosing an unexpected event. In line with this hypothesis pilots spent a greater amount of time looking at the flight displays and its various controls following the stall horn in the unexpected stall. This is alarming as there should be an automatic association between the stall horn, the diagnosis and the action selection. However this information processing pattern displayed by pilots in the unexpected stall flight provides evidence suggesting that this may not be the case. Future research should focus on establishing the effectiveness of the stall horn, or the stick shaker as indicators of stall, and their ability to induce pilots to implement the correct recovery procedures. This is especially important in off-normal stall events which are more relevant to real-life emergencies in comparison to the practice stall procedure.

***Does physiological startle impair response performance following unexpected emergency events?***

There were no significant differences in heart rate or pupil dilation for the pilots or students that crashed in the unexpected engine failure compared those that landed safely. In fact during the unexpected engine failure, those pilots that landed safely had a higher heart rate and larger pupil dilation than those that crashed. This is the inverse relationship that was expected. Additionally, pilots who responded to the stall horn correctly in the unexpected stall showed no significant differences in heart rate or pupil dilation compared to those that did not respond correctly.

Recent researchers have postulated that startle, due to its associated cognitive impairment, is a potential factor in accident causation (Landman et al., 2017a, 2017b; Martin et al., 2015; Martin et al., 2010; Martin et al., 2016; Schroeder et al., 2014). Impairment in the present study was not associated with any significant differences in heart rate or pupil dilation. However this between-subjects analysis was limited by small sample size leading to small power. Further research with larger samples needs to be completed to investigate this research question.

Research on fear conditioning has shown that when startle occurs in the presence of perceived threat the response can become exacerbated leading to what is known as fear potentiated startle (Bradley, Moulder, & Lang, 2005; Eysenck et al., 2005). As there was no real threat to life in the flight simulator, pilots startle reaction is unlikely to have fully escalated into fear-potentiated startle. This may be why there were no significant differences in autonomic arousal between participants that performed well and those that performed poorly. Due to this limitation, it cannot be discounted that fear-potentiated startle could be impacting performance in abnormal flight events.

Unfortunately the effects of fear-potentiated startle be very difficult to investigate. It may be possible to research performance and arousal in a high-fidelity simulator with airline pilots who are undergoing their six monthly flight simulator testing. This flight testing determines whether they can continue to fly or need further training. Even though there is no threat to the pilot’s lives, there is threat to their livelihood. In accordance with Landman et al.’s (2017a) conceptual model of startle, perception of an abnormal event combined with threat leads to startle (Figure 1). The threat of failure during the six month simulator testing for commercial pilots could be threatening enough to lead to fear-potentiated startle in response to an unexpected emergency event. Therefore, it may be a promising idea to capitalise on this situation by completing a similar study to the present research in an operational testing environment.

***Does expertise mediate the startle response during unexpected emergency flight events?***

The results from the present study suggest that expertise does not mediate the startle reaction in unexpected emergency events. The present research failed to find any significant differences in heart rate or pupil dilation for the expert pilots and the intermediate pilots for any of the flights. University students had the lowest heart rate in the unexpected engine failure, the expected engine failure, and baseline. There were no significance differences in pupil dilation between the university students and both the pilot groups for the unexpected engine failure and stall flights or the expected stall flight. This is consistent with Casner et al. (2012) and Martin et al. (2016) who found that there was no significant effect of increased flight hours on performance in unexpected simulated tasks.

The low level of arousal for university students may have been due to lack of motivation to perform well, (Bergman & Magnusson, 1979; Lazarus et al., 1952; Skinner & Brewer, 2002; Vogel et al., 1959) or due to low levels of immersion in the simulator (Alexander et al., 2005; Brown & Cairns, 2004; Cairns et al., 2014; Cheng & Cairns, 2005; Sadowski & Stanney, 2002; Witmer & Singer, 1998). However surprisingly university students had the highest increase in pupil dilation in the expected engine failure compared to the intermediate pilots. This could be due to higher cognitive workload for the students, as pilots were relatively more experienced at landing.

Establishing whether experience mediates the startle effect is very important in providing evidence for the viability of the current training methods. There were non-significant differences in the mean heart rates over the flights between expert pilots and intermediate pilots. Intermediate pilots had higher heart rates in all events. These may be true differences that were not significant due to a lack of power in the between-subjects analyses. Therefore, it would be interesting to complete a between-subjects analysis with at least 20 intermediate level general aviation pilots, and 20 expert level general aviation pilots.

***Does expertise affect information processing differentially during unexpected emergency flight events?***

Contrary to the hypotheses there were no significant differences in the amount of time each group spent viewing the flight displays in the unexpected engine failure. During the unexpected stall there was a significant overall effect of expertise on time spent viewing the airspeed indicator, however post hoc tests showed no significant differences between individual groups. There were detectable slight differences in the information processing strategies used by novices compared to pilots overall in the unexpected abnormal flight events. Speculation on the non-significant data suggest that there may have been real differences in the information processing over the two different pilot groups. However, this present research cannot make any definitive conclusions.

**Limitations**

The present study was both challenging, and ambitious. However, it also has a few limitations. As previously mentioned the between-subject analyses and the stall analyses were limited by a small sample size. Due to the stall horn only sounding in around 60% of pilot's flights, the sample size for the stall analyses was limited, which probably led to low power. Also smaller sample sizes when separating the pilots into different groups for between-subjects analyses also led to low power. However, when comparing with other research in the human factors field, twenty-one pilots is a reasonably good sample. However these limited samples have likely restricted some of the findings, where some of the non-significant findings may indicate real differences. Therefore, researchers with good access to pilots could repeat the between-subjects analyses completed in this study with a larger sample size, or with a more sophisticated simulator where the coded stall means pilots have a 100% chance of encountering the stall horn.

Another limitation is that the performance and information processing in the emergency events may have been affected by the fidelity of the simulator. Koonce (1984) found that predictors of performance were emotional stability, perception of tilt, and mental alertness  In the stall flight students and pilots were unable to perceive the tilt of the aircraft this could have affected performance, especially in the stall flight. As well as this, in the unexpected engine failure pilots would have been very limited by not being able to look out the windows for a 360° view on potential landing spots. Future research in higher fidelity simulators would be beneficial in providing evidence supporting the findings of the present research.

Furthermore, in hindsight the conclusions and findings, the study may have benefitted from having a questionnaire completed at the end of the simulator session or at the end of each of the experimental flights. It could have been beneficial for the clarity of the study to have a subjective measure of workload and/or startle in the experimental flights. As well as this, in the unexpected stall event flight. It may have been helpful to have a subjective measure of recognition and the time to recognize the event. The subjective measures of recognition could have supported the conclusions made in the information processing analyses. Specifically that the pilots spent more time looking at the flight displays as they were diagnosing the unexpected stall. A subjective measure of workload and startle would have been able to directly support the conclusions made regarding physiological arousal. However, addition of these questionnaires could have been disruptive to the flight task, and could have limited immersion in the task. As well as this, the validity of subjective questionnaire answers could have been compounded by the outcome of the flight. Future research should make sure they have both subjective and objective measures of the key dependent variables, however not at the expense of the quality of the experiment.

**Future Research**

Opportunities for future research have been indicated throughout the thesis and the final chapter. The opportunities indicated in the other chapters are summarised in the following bullet points:

- Extend the findings of the present study using a high-fidelity simulator and/or commercial pilots.
- Further research with an experimental design providing increased power to find differences when investigating the concept that high arousal analogous to startle leads to impaired performance.
- Further research with an experimental design providing increased power to find differences when investigating the concept that expertise may mediate the negative impact of startle in unexpected flight events.
- Investigation into the differential responses to abnormal events in accordance with Landman et al.'s (2017a) model. In particular, specific characterisation of the different main incorrect responses to abnormal events.
- Another possibility of studying expertise would be to have novice pilots practice flying and emergency events over time and until proficiency before encountering the unexpected versions of events.

*New research idea:* Automation is thought to degrade situational awareness (see Endsley (1999) for a review). There is another research opportunity in regards to automation and unexpected events. It would be very interesting to research startle, expertise, and information processing during an unexpected event that has occurred after a prolonged period of partial automation.

**Conclusion**

The findings of the current study support the idea that unexpected events in a flight simulator can lead to physiological arousal that is consistent with startle. If the abnormal event is easily recognized and threatening (e.g. engine failure) information processing can be impaired by attentional tunnelling where less time is spent viewing the flight display and critical instruments. If the abnormal event is not easily recognized, pilots tend to spend more time viewing the flight display and the critical instruments. This could also be viewed as an impairment, as in abnormal events where the event should be instantly recognized, pilots may be losing critical time diagnosing the situation. There were no significant differences in autonomic arousal for pilots that performed well and those that performed poorly in the different event types. Furthermore, there was no significant effect of expertise on autonomic arousal or information processing. However the non-significant data suggest that there are possibly true differences that have not emerged due to lack of power. The present research supports Landman et al's (2017a) conceptual model of startle and surprise. It also supports Brown and Cairns' (2005) theory of the three levels of immersion where novices were not affected emotionally by the unexpected emergency events as indicated by their lack of autonomic arousal following the commencement of the abnormal events. There are many opportunities for future research in this field, and it is certain the effects of startle and surprise in unexpected events could be extrapolated to many other human factors fields such as surgery, rail operation, driver safety, and other technological industries.

# References


Aeronautical Information Publication New Zealand (2002, June 2017). *Aerodrome Charts.* Retrieved from http://www.aip.net.nz/navwalk.aspx?section=charts.

Advani, S. K., Schroeder, J. A., & Burks, B. (2010). *What Really Can Be Done in Simulation to Improve Upset Training?* Paper presented at the Proceedings of the AIAA Guidance, Navigation, and Control Conference.

Agelink, M. W., Malessa, R., Baumann, B., Majewski, T., Akila, F., Zeit, T., & Ziegler, D. (2001). Standardized tests of heart rate variability: Normal ranges obtained from 309 healthy humans, and effects of age, gender, and heart rate. *Clinical Autonomic Research, 11*(2), 99-108. doi: 10.1007/bf02322053

Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group, 5*, 1-14.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829.

Baddeley, A. D. (1972). Selective attention and performance in dangerous environments. *British Journal of Psychology, 63*(4), 537-546.

Barnes, V. E., & Monan, W. P. (1990). Cockpit Distractions: Precursors to Emergencies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 34*(16), 1142-1144. doi: 10.1177/154193129003401610

Basner, M. (2009). Pilot workload during approaches: Comparison of simulated standard and noise-abatement profiles. *Aviation, 80*(4), 364-370. doi: 10.3357/ASEM.2382.2009

Belcastro, C. M., & Foster, J. V. (2010). *Aircraft loss-of-control accident analysis.* Paper presented at the Proceedings of AIAA Guidance, Navigation and Control Conference, Toronto, Canada, Paper No. AIAA-2010-8004.

Bergman, L. R., & Magnusson, D. (1979). Overachievement and catecholamine excretion in an achievement-demanding situation. *Psychosomatic Medicine*. *51*(3), 181-188.

Boot, W. R., & Ericsson, K. A. (2013). Expertise. In J. D. Lee, & A. Kirlik (Eds.), *The Oxford Handbook of Cognitive Engineering* (pp. 143–158). New York: Oxford University Press

Bourne Jr, L. E., & Yaroush, R. A. (2003). Stress and cognition: A cognitive psychological perspective.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607. doi: 10.1111/j.1469-8986.2008.00654.x

Bradley, M. M., Moulder, B., & Lang, P. J. (2005). When good things go bad: The reflex physiology of defense. *Psychological Science, 16*(6), 468-473.

Broadbent, D. E. (1971). *Decision and stress*. Oxford, England: Academic Press. http://dx.doi.org/.

Brosschot, J. F., & Thayer, J. F. (2003). Heart rate response is longer after negative emotions than after positive emotions. *International Journal of Psychophysiology, 50*(3), 181-187.

Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. *In CHI'04 extended abstracts on Human factors in computing systems* (pp. 1297-1300). ACM.

Bürki-Cohen, J. (2010). *Technical challenges of upset recovery training: Simulating the element of surprise.* Paper presented at the Proceedings of the AIAA Guidance, Navigation, and Control Conference.

Burlingame Software. (2009). Flight Data Recorder. Retrieved November 1, 2016, from http://www.software.burlingames.com/recorder.php#

Cairns, P., Cox, A., & Nordin, A. I. (2014). Immersion in digital games: review of gaming experience research. *Handbook of Digital Games, 1*, 767.

Carlsen, A. N., Chua, R., Dakin, C. J., Sanderson, D. J., Inglis, J. T., & Franks, I. M. (2008). Startle reveals an absence of advance motor programming in a Go/No-go task. *Neuroscience Letters, 434*(1), 61-65.

Casner, S. M., Geven, R. W., & Williams, K. T. (2012). The effectiveness of airline pilot training for abnormal events. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720812466893.

Cheng, K., & Cairns, P. A. (2005, April). Behaviour, realism and immersion in games. *In CHI'05 extended abstracts on Human factors in computing systems* (pp. 1272-1275). ACM.

Chou, C.-Y., Marca, R. L., Steptoe, A., & Brewin, C. R. (2014). Heart rate, startle response, and intrusive trauma memories. *Psychophysiology, 51*(3), 236-246. doi: 10.1111/psyp.12176

Civil Aviation Authority, N. Z. (2012). Flight Instructor Guide. Retrieved from https://www.caa.govt.nz/fig/index.html

Combs, A. W., & Taylor, C. (1952). The effect of the perception of mild degrees of threat on performance. *The Journal of Abnormal and Social Psychology, 47*(2S), 420.

Davis, M. (1984). The mammalian startle response *Neural mechanisms of startle behavior* (pp. 287-351): Springer, Boston, MA.

Deuter, C. E., Kuehl, L. K., Blumenthal, T. D., Schulz, A., Oitzl, M. S., & Schachinger, H. (2012). Effects of cold pressor stress on the human startle response. *PloS one, 7*(11), e49866.

Dismukes, R. K., Goldsmith, T. E., & Kochan, J. A. (2015). Effects of acute stress on aircrew performance: Literature review and analysis of operational aspects: NASA Report: NASA/TM-2015–218930. Moffett Field, CA: National Aeronautics and Space Administration.

Driskell, J. E., & Salas, E. (2013). *Stress and human performance*: Psychology Press.

Duncko, R., Johnson, L., Merikangas, K., & Grillon, C. (2009). Working memory performance after acute exposure to the cold pressor stress in healthy volunteers. *Neurobiology of Learning and Memory, 91*(4), 377-381.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological review, 66*(3), 183.

Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences, 105*(5), 1704-1709.

Endsley, M. R. (1999). Situation awareness in aviation systems. *Handbook of aviation human factors*, 257-276.

Eysenck, M., Payne, S., & Derakshan, N. (2005). Trait anxiety, visuospatial processing, and working memory. *Cognition & Emotion, 19*(8), 1214-1228.

Federal Aviation Administration. (1988). *Pilot Windshear Guide*. (00-5). Washington, D.C: U.S. G.P.O.

Federal Aviation Administration. (2015). Upset prevention and recovery training (Advisory Circular AC 120-111). Retrieved from http://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-111.pdf

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., & Miyake, S. (2016). Assessment of operators' mental workload using physiological and subjective measures in cement, city traffic and power plant control centers. *Health promotion perspectives, 6*(2), 96.

Froelicher, V. F., & Myers, J. (2007). Chapter 3 - Interpretation of Hemodynamic Responses to Exercise Testing *Manual of Exercise Testing (Third Edition)* (pp. 51-85). Philadelphia: Mosby.

Godzone Virtual Flight. (n.d.). In WindowLight. Retrieved September 15, 2016, from http://windowlight.co.nz/store/

Gomes, L. M., Martinho Pimenta, A. J., & Castelo Branco, N. A. (1999). Effects of occupational exposure to low frequency noise on cognition. *Aviat Space Environ Med, 70*(3 Pt 2), A115-118.

Grassmann, M., Vlemincx, E., Von Leupoldt, A., & Van Den Bergh, O. (2017). Individual differences in cardiorespiratory measures of mental workload: An investigation of negative affectivity and cognitive avoidant coping in pilot candidates. *Applied Ergonomics, 59*, 274-282. doi: 10.1016/j.apergo.2016.09.006

Green, R. (1985). Stress and accidents. *Aviation, Space, and Environmental Medicine*.

Hodges, W., & Spielberger, C. (1966). The effects of threat of shock on heart rate for subjects who differ in manifest anxiety and fear of shock. *Psychophysiology, 2*(4), 287-294.

Holand, S., Girard, A., Laude, D., Meyer-Bisch, C., & Elghozi, J.-L. (1999). Effects of an auditory startle stimulus on blood pressure and heart rate in humans. *Journal of Hypertension, 17*(12), 1893-1897.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology Section A, 48*(3), 598-612.

Jacobson, S. R. (2010). Aircraft loss of control causal factors and mitigation challenges. *American Institute of Aeronautics and Astronautics, 8007*, 2-5.

Jensen, R. S. (1995). *Pilot judgement and crew resource management*. Aldershot, Hants ; Brookfield, Vt.: Aldershot, Hants ; Brookfield, Vt. : Avebury Aviation.

Jiang, C., & Rau, P.-L. P. (2017). Working memory performance impaired after exposure to acute social stress: The evidence comes from ERPs. *Neuroscience Letters, 658*, 137-141. doi: https://doi.org/10.1016/j.neulet.2017.08.054

Kahneman, D. (1973). *Attention and effort* (Vol. 1063): Prentice-Hall Englewood Cliffs, NJ.

King, R., & Schaefer, A. (2011). The emotional startle effect is disrupted by a concurrent working memory task. *Psychophysiology*, *48*(2), 269-272.

Kirschenbaum, S. S. (1992). Influence of experience on information-gathering strategies. *Journal of Applied Psychology, 77*(3), 343.

Kivimäki, M., & Lusa, S. (1994). Stress and cognitive performance of fire fighters during smoke-diving. *Stress and Health, 10*(1), 63-68.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). *A data-frame theory of sensemaking.* Paper presented at the Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making.

Koonce, J. M. (1984). A Brief History of Aviation Psychology. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 26*(5), 499-508. doi: 10.1177/001872088402600502

Lahtinen, T. M., Koskelo, J. P., Laitinen, T., & Leino, T. K. (2007). Heart rate and performance during combat missions in a flight simulator. *Aviation, Space, and Environmental Medicine, 78*(4), 387-391.

Landman, A., Groen, E. L., van Paassen, M., Bronkhorst, A. W., & Mulder, M. (2017a). Dealing with unexpected events on the flight deck: a conceptual model of startle and surprise. *Human Factors, 59*(8), 1161-1172.

Landman, A., Groen, E. L., Van Paassen, M., Bronkhorst, A. W., & Mulder, M. (2017b). The influence of surprise on upset recovery performance in airline pilots. *The International Journal of Aerospace Psychology, 27*(1-2), 2-14.

Lazarus, R. S., Deese, J., & Osler, S. F. (1952). The effects of psychological stress upon performance. *Psychological Bulletin, 49*(4), 293.

Ledegang, W. D., & Groen, E. L. (2015). Stall recovery in a centrifuge-based flight simulator with an extended aerodynamic model. *The International Journal of Aviation Psychology, 25*(2), 122-140.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology, 23*(4-5), 727-738.

Lehner, P., Seyed-Solorforough, M. M., Connor, M. F. O., Sak, S., & Mullin, T. (1997). Cognitive biases and time stress in team decision making. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 27*(5), 698-703. doi: 10.1109/3468.618269

Li, G., Baker, S. P., Grabowski, J. G., & Rebok, G. W. (2001). Factors associated with pilot error in aviation crashes. *Aviation, Space, and Environmental Medicine, 72*(1), 52-58.

Loveday, T., Wiggins, M., Festa, M., Schell, D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In *Pattern recognition-Applications and methods*(pp. 1-11). Springer, Berlin, Heidelberg.

Main, L. C., Wolkow, A., & Chambers, T. P. (2017). Quantifying the Physiological Stress Response to Simulated Maritime Pilotage Tasks: The Influence of Task Complexity and Pilot Experience. *Journal of Occupational and Environmental Medicine, 59*(11), 1078.

Mannaru, P., Balasingam, B., Pattipati, K., Sibley, C., & Coyne, J. T. (2017). Performance Evaluation of the Gazepoint GP3 Eye Tracking Device Based on Pupil Dilation. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Augmented Cognition. Neurocognition and Machine Learning: 11th International Conference, AC 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I* (pp. 166-175). Cham: Springer International Publishing.

Marinescu, A. C., Sharples, S., Ritchie, A. C., Sánchez López, T., McDowell, M., & Morvan, H. P. (2018). Physiological Parameter Response to Variation of Mental Workload. *Human factors, 60*(1), 31-56. doi: 10.1177/0018720817733101

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *Human factors and power plants, 2002. proceedings of the 2002 IEEE 7th conference on* (pp. 7-7). IEEE.

Martin, W. L., Bates, P. R., & Murray, P. S. (2010). The effects of stress on pilot reactions to unexpected, novel, and emergency events. In *Proceedings of the 9th International Australian Aviation Psychology Association Symposium* (pp. 18-22).

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. (2015). Fear-Potentiated Startle: A Review from an Aviation Perspective. *The International Journal of Aviation Psychology, 25*(2), 97-107.

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. (2016). A Flight Simulator Study of the Impairment Effects of Startle on Pilots During Unexpected Critical Events. *Aviation Psychology and Applied Human Factors*, *6*(1), 24.

May, D., & Rice, C. (1971). Effects of startle due to pistol shots on control precision performance. *Journal of Sound and Vibration, 15*(2), 197IN1199-198202.

McKinney, E. H. (1993). Flight leads and crisis decision-making. *Aviation, Space, and Environmental Medicine*. 64, 359-362.

McKinney Jr, E. H., & Davis, K. J. (2003). Effects of deliberate practice on crisis decision performance. *Human factors, 45*(3), 436-444.

Melton, A. (1947). Apparatus tests: Army Air Forces Aviation Psychology Program Research Reports Number 4. *Washington, DC: US Government Printing Office*.

Nevile, M. (2001). Understanding who's who in the airline cockpit: pilots' pronominal choices and cockpit roles. *How to analyse talk in institutional settings: A casebook of methods*, 57-71.

National Transportation Safety Board. (1995). *Aircraft accident report: Flight into terrain during missed approach, USAir Flight 1016, Charlotte, North Carolina, July 2, 1994*. Washington, DC: Author.

National Transportation Safety Board. (2004). *Aircraft accident report: In-Flight Engine Failure and Subsequent Ditching Air Sunshine, Inc., Bahamas, September 13, 2003*. Washington, SC: Author.

National Transportation Safety Board. (2010a). *Aircraft accident report: Loss of control on approach, Colgan Air, Inc. Clarence Center, New York, February 12, 2009*. Washington, DC: Author.

National Transportation Safety Board. (2010b). *Aircraft accident report: Runway overrun during rejected takeoff, Global Exec Aviation, Columbia, South Carolina, September 19, 2008*. Washington, DC: Author.

O'Hare, D., Wiggins, M., Batt, R., & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics, 37*(11), 1855-1869.

O'Hare, D. (2003). Aeronautical decision making: Metaphors, models, and methods. *Principles and Practice of Aviation Psychology*, 201-237.

Orbx Simulation Systems (2017) NZ South Island. Retrieved April 28, 2017, from https://orbxdirect.com/product/nzsi

Orr, S. P., Solomon, Z., Peri, T., Pitman, R. K., & Shalev, A. Y. (1997). Physiologic responses to loud tones in israeli veterans of the 1973 yom kippur war. *Biological Psychiatry, 41*(3), 319-326. doi: https://doi.org/10.1016/S0006-3223(95)00671-0

Porcelli, A. J., Cruz, D., Wenberg, K., Patterson, M. D., Biswal, B. B., & Rypma, B. (2008). The effects of acute stress on human prefrontal working memory systems. *Physiology & Behavior, 95*(3), 282-289.

Proctor, R. W., & Van Zandt, T. (2018). *Human factors in simple and complex systems*: CRC press. Needham Heights, MA: AUyn & Bacon.

Raney, G. E., Campbell, S. J., & Bovee, J. C. (2014). Using Eye Movements to Evaluate the Cognitive Processes Involved in Text Comprehension. *Journal of Visualized Experiments : JoVE*(83), 50780. doi: 10.3791/50780

Raschke, M., Blascheck, T., & Burch, M. (2014). Visual analysis of eye tracking data. In *Handbook of Human Centric Visualization* (pp. 391-409). Springer, New York, NY.

Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014). *Startle and Surprise on the Flight Deck Similarities, Differences, and Prevalence.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Sadowski, W., & Stanney, K. (2002). Presence in virtual environments. In K. M. Stanney (Ed.), *Human factors and ergonomics. Handbook of virtual environments: Design, implementation, and applications* (pp. 791-806). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8*(3), 197-208.

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols.* Paper presented at the Proceedings of the 2000 symposium on Eye tracking research & applications.

Schroeder, J. A., Burki-Cohen, J. S., Shikany, D., Gingras, D. R., & Desrochers, P. P. (2014). *An evaluation of several stall models for commercial transport training.* Paper presented at the AIAA Modeling and Simulation Technologies Conference.

Skinner, N., & Brewer, N. (2002). The dynamics of threat and challenge appraisals prior to stressful achievement events. *Journal of Personality and Social Psychology, 83*(3), 678.

Staal, M. A. (2004). Stress, cognition, and human performance: A literature review and conceptual framework. Hanover, MD: National Aeronautics & Space Administration.

Sternbach, R. A. (1960). Correlates of differences in time to recover from startle. *Psychosomatic Medicine, 22*(2), 143-148.

Stokes, A., Belger, A., & Zhang, K. (1990). Investigation of factors comprising a model of pilot decision making, part II: Anxiety and cognitive strategies in expert and novice aviators. *University of Illinois Aviation Research Laboratory, Savoy*.

Streufert, S., & Streufert, S. C. (1981). Stress and information search in complex decision making: Effects of load and time urgency: Milton S Hershey Medical Center Pa Dept Of Behavioral Science.

Thackray, R. I. (1965). Correlates of Reaction Time to Startle. *Human Factors, 7*(1), 74-80. doi: 10.1177/001872086500700109

Thackray, R. I. (1988). Performance recovery following startle: A laboratory approach to the study of behavioral response to sudden aircraft emergencies: Federal Aviation Administration Washington Dc Office Of Aviation Medicine.

Thackray, R. I., & Touchstone, R. M. (1970). Recovery of Motor Performance following Startle. *Perceptual and Motor Skills, 30*(1), 279-292. doi: 10.2466/pms.1970.30.1.279

Thackray, R. I., & Touchstone, R. M. (1983). Rate of initial recovery and subsequent radar monitoring performance following a simulated emergency involving startle: Federal Aviation Administration, Oklahoma City, Oklahoma.

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2): Reading, Mass.

Urbaniak, G. C., & Plous, S. (1997, 5/05/2017). Research randomizer. *URL: https://www.randomizer.org/*

Vaidya, A. R., & Fellows, L. K. (2017). Chapter 22 - The Neuropsychology of Decision-Making: A View From the Frontal Lobes *Decision Neuroscience* (pp. 277-289). San Diego: Academic Press.

Vlasak, M. (1969). Effect of startle stimuli on performance. *Aerospace medicine, 40*(2), 124.

Vogel, W., Raymond, S., & Lazarus, R. S. (1959). Intrinsic motivation and psychological stress. *The Journal of Abnormal and Social Psychology, 58*(2), 225.

Wickens, C. D., Stokes, A., Barnett, B., & Hyman, F. (1993). The effects of stress on pilot judgment in a MIDIS simulator. In *Time Pressure and Stress in Human Judgment and Decision Making* (pp. 271-292). Springer, Boston, MA.

Wiggins, M., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied, 1*(4), 305.

Wiggins, M., Stevens, C., Howard, A., Henley, I., & O'Hare, D. (2002). Expert, intermediate and novice performance during simulated pre-flight decision-making. *Australian Journal of Psychology, 54*(3), 162-167.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225-240.

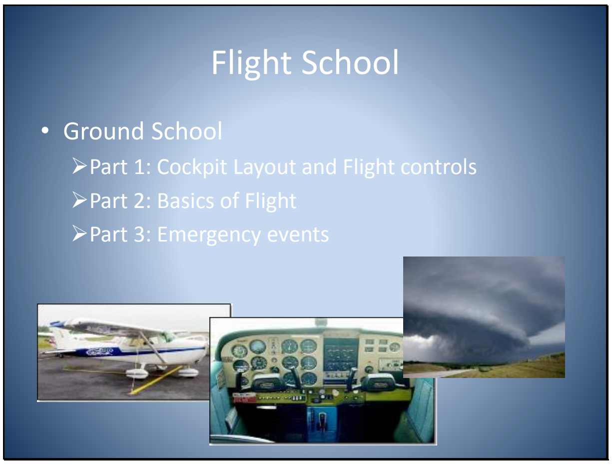
Woodhead, M. M. (1969). Performing a visual task in the vicinity of reproduced sonic bangs. *Journal of Sound and Vibration, 9*(1), 121-125. doi: https://doi.org/10.1016/0022-460X(69)90269-7

Woodson, W. E., Tillman, B., & Tillman, P. (1992). *Human Factors Design Handbook: Information and guidelines for the design of systems, facilities, equipment, and products for human use*. New York:McGraw-Hill, 1992.

Appendix A. Flight School Powerpoint

Slide 1

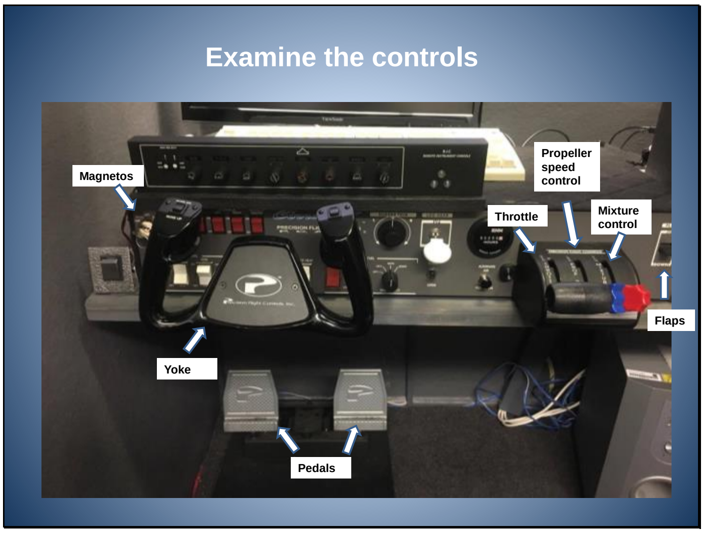

Slide 2




Slide 3

Slide 4

Slide 5

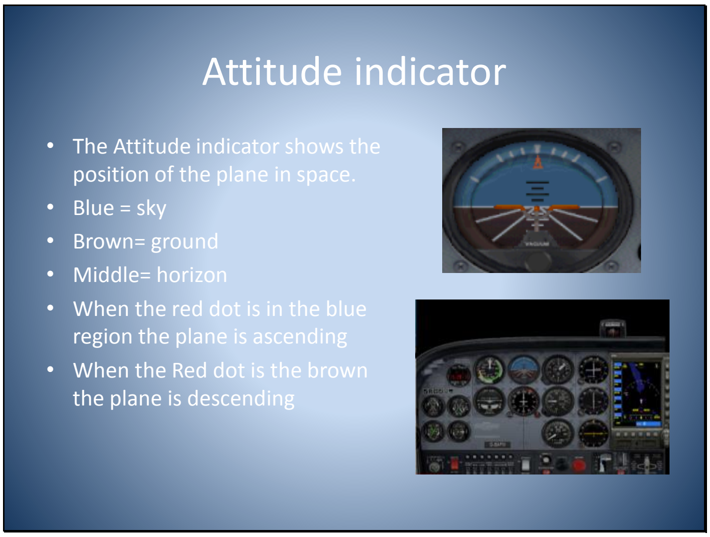



Slide 6

Slide 7

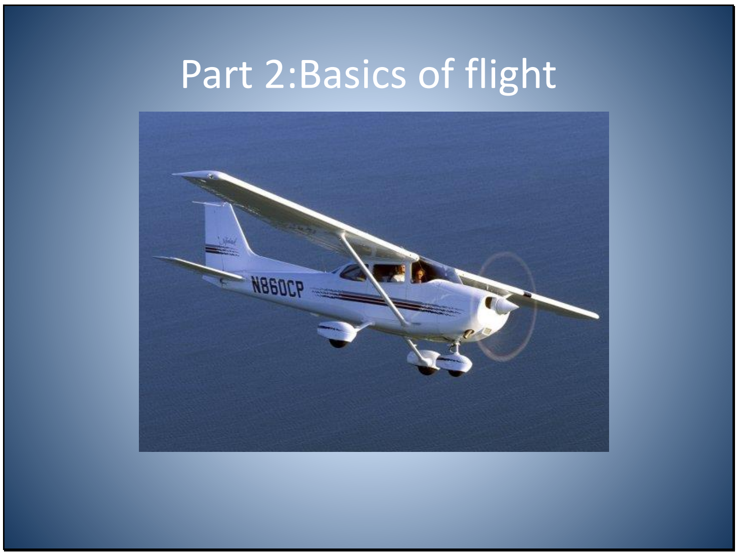

Slide 8


Slide 9



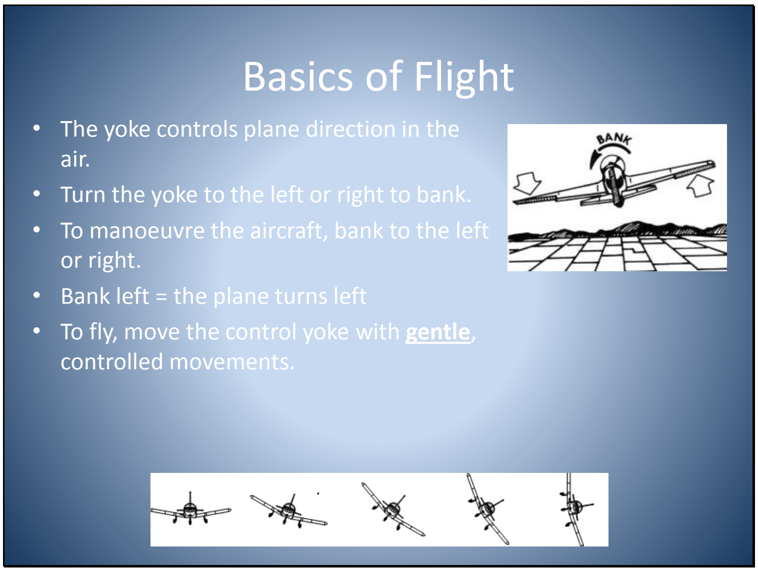
Slide 10


Slide 11

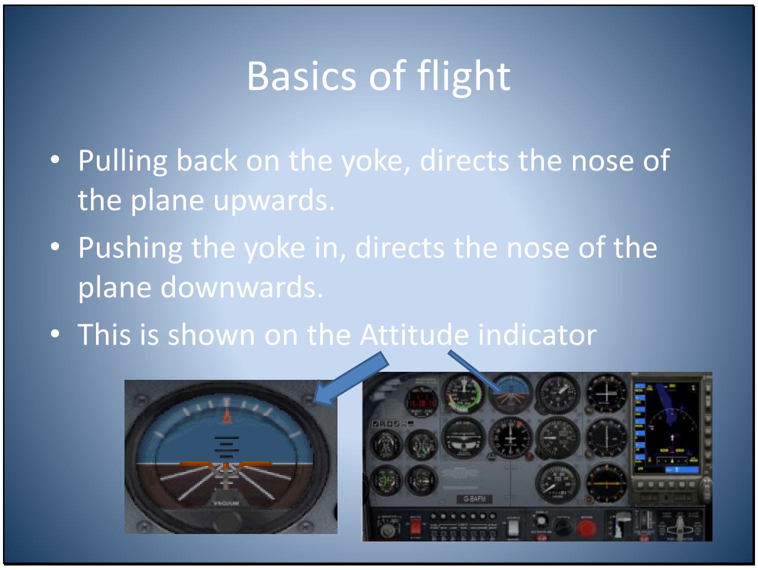



Slide 12

Slide 13




Slide 14



Slide 15

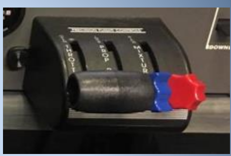
## Basics of Flight

- Remember! Pointing the nose above the horizon (going up) or below the horizon (going down) affects your speed.
- Push forward (nose down) to go faster, and pull back (nose up) to slow down.
- When flying completely level, adjusting the throttle allows you to climb and descend.

Slide 16

## Basics of Flight :To turn on the plane

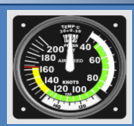
- Push the blue (propeller speed) and red (mixture control) levers up.
- Turn the left magneto from OFF to BOTH.
- Push the Engine start switch to the left.



Slide 17

## Basics of Flight:To take off

- Make sure the parking brake is released. If parking brakes are on, it will tell you at the bottom of the screen.
- Push the flaps control down once to get 10 degrees of flaps.
- Increase the throttle to full.
- Use the pedals to direct the plane straight down the runway
- When you reach a airspeed of at least 60 knots, pull back on the yoke to ascend.

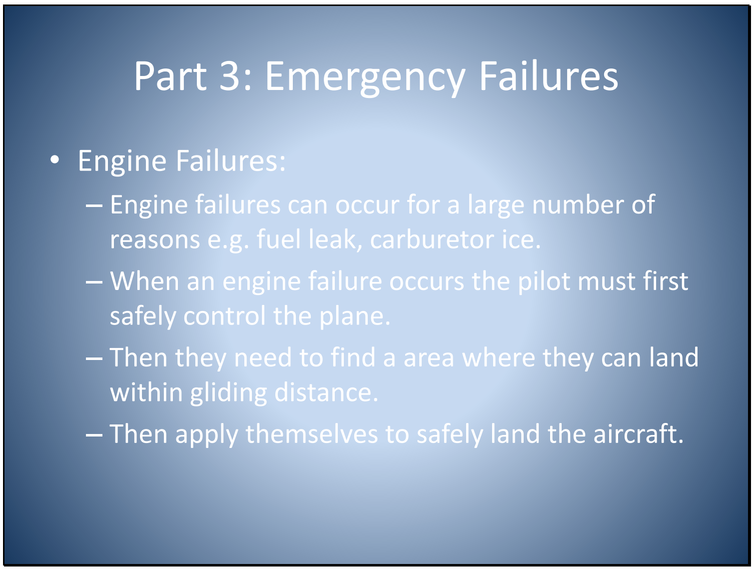


Airspeed Indicator

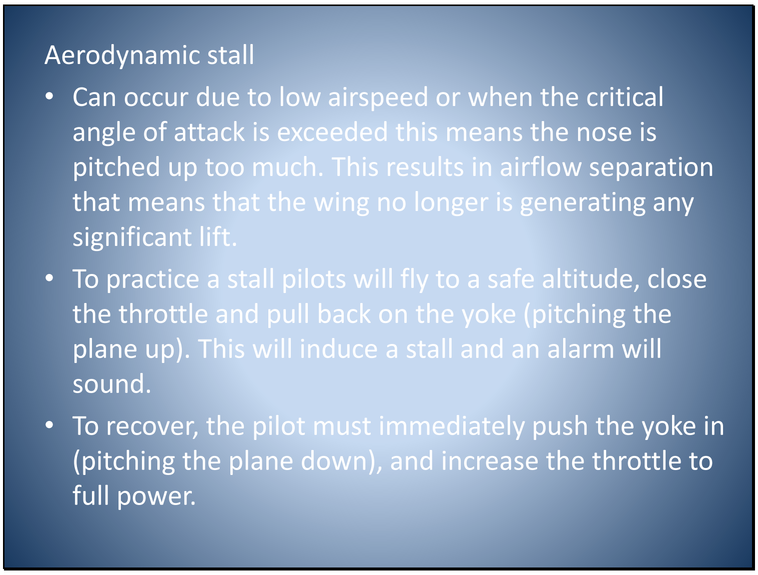
Displays current airspeed in knots (nautical miles per hour)

Slide 18


Slide 19




Slide 20



Slide 21

Appendix B. University Student Questionnaire

# Flight School: A Flight Simulator Experiment

***Demographics and Experience Questionnaire***

1. **How old are you:**
2. **What is your gender (circle one):** Male Female Neutral
3. **Do you have previous flight experience (circle one)?**
    a. Yes – I have flown an aircraft before.
    b. Yes - I have flown a flight simulator before.
    c. No – this will be my first time flying a plane.

    If you answered '**Yes'** please describe the amount of flight/flight simulator experience you have:
    ______________________________________________________________________
    ______________________________________________________________________
    ___________________________________________________________
4. **How much experience with computer or console gaming would you identify as having (circle one)?**
    a. Gaming Expert (You play every day if possible).
    b. Social Gamer (You play once or twice a week).
    c. Infrequent Gamer (You might play occasionally).
    d. Not a Gamer (You very rarely play games).
    e. Other (please leave notes)___________________________________________________________
    ________________________________________________________________
    ______________________________________________________
5. **Regardless of your current gaming levels have you ever identified as a gamer in the past (circle one)?**
    a. Yes (please give some details)
    ________________________________________________________________
    ________________________________________________________________
    ________________________________________.
    b. No

Appendix C. Information Sheet for Student Participants

21/17
27/08/2017

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

***Pilot reactions to differential flight conditions.***
**INFORMATION SHEET FOR PARTICIPANTS**

Thank you for showing an interest in this project. Please read this information sheet carefully before deciding whether or not to participate. If you decide to participate we thank you. If you decide not to take part there will be no disadvantage to you and we thank you for considering our request.

**What is the Aim of the Project?**

This is an experiment investigating pilot reactions to different flight conditions. You will be asked to fly five small flights around New Zealand in our Microsoft flight simulator. We are investigating eye movements, heart rate, and flight performance in different flight conditions including training events e.g. power-off stalls. This project is being undertaken as part of the requirements for Lana Kinney's Masters of Science. This study will help improve the understanding of pilot behavioural reactions to in-flight events.

**What Types of Participants are being sought?**

We are seeking at least 30 100 level student participants to be trained on the flight simulator and complete the flights. Students must either have good vision or wear contacts as glasses can unfortunately interfere with the eye tracking software. Participants will be asked whether they would like a copy of the experimental findings when the project is finished.

**What will Participants be asked to do?**

Should you agree to take part in this project, you will be asked to undergo flight training and then fly approximately 8 short flights around New Zealand in the psychology department's flight simulator. The experiment will take around 2 hours. You will be first asked to fill out a short demographics questionnaire, then you will watch a flight training PowerPoint. During each flight heart rate will be recorded with an unobtrusive ear clip, eye movements will be remotely recorded, and flight performance data will be automatically logged.

Please be aware that you may decide not to take part in the project without any disadvantage to yourself.

**What Data or Information will be collected and what use will be made of it?**

Your eye movements, heart rate, and flight data during the flights will be collected and stored electronically. The questionnaire will ask for your name, age, and flight experience information. The data will be used to examine your responses to different flight conditions. Apart from your name, and age, no other personal information will be collected. The researchers below will have access to the data. Results from this project may be published and will be available in the University of Otago Library (Dunedin, New Zealand), but every attempt will be made to preserve your anonymity. You will not be identifiable in the publications. The data collected will be securely stored in such a way that only those mentioned below will be able to gain access to it. At the end of the project any personal information will be destroyed immediately except that, as required by the University's research policy, any raw data on which the results of the project depend on will be retained in secure storage for at least five years, after which it will be destroyed. You can choose to be provided with a summary of the study's results if you wish.

**Can Participants change their mind and withdraw from the project?**

You may withdraw from participation in the project at any time during the experiment and without any disadvantage to yourself.

**What if Participants have any Questions?**

If you have any questions about our project, either now or in the future, please feel free to contact either:-

*Lana Kinney* or *David O'Hare*

| | |
|---|---|
| *Lana Kinney* | *David O'Hare* |
| Department of Psychology | Department of Psychology |
| Telephone Number: 027 399 7122 | Telephone Number: (64)3-479-7643 |
| Email Address: lanakinney92@gmail.com | Email Address: ohare@psy.otago.ac.nz |

This study has been approved by the Department stated above. However, if you have any concerns about the ethical conduct of the research you may contact the University of Otago Human Ethics Committee through the Human Ethics Committee Administrator (ph 03 479-8256). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome.

Appendix D. Consent Form for Student Participants

***PILOT REACTIONS TO DIFFERENTIAL FLIGHT CONDITIONS.***

***CONSENT FORM FOR***

***PARTICIPANTS***

I have read the Information Sheet concerning this project and understand what it is about. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage.

I know that:-

1. My participation in the project is entirely voluntary;

2. I am free to withdraw from the project at any time without any disadvantage;

3. Personal identifying information (name and age) will be destroyed at the conclusion of the project but any raw data on which the results of the project depend will be retained in secure storage for at least five years.

4. There is no foreseeable discomfort or distress for the participant while taking part in this study

5. The results of the project may be published and will be available in the University of Otago Library (Dunedin, New Zealand) but every attempt will be made to preserve my anonymity.

I agree to take part in this project.

............................................................................. ...............................

(Signature of participant) (Date)

.............................................................................

(Printed Name)

Appendix E. Information Sheet for Pilot Participants

***Pilot reactions to differential flight conditions.***
**INFORMATION SHEET FOR PARTICIPANTS**

Thank you for showing an interest in this project. Please read this information sheet carefully before deciding whether or not to participate. If you decide to participate we thank you. If you decide not to take part there will be no disadvantage to you and we thank you for considering our request.

**What is the Aim of the Project?**

This is an experiment investigating pilot reactions to different flight conditions. You will be asked to fly eight small flights around New Zealand in our Microsoft flight simulator. We are investigating eye movements, heart rate, and flight performance in different flight conditions including training events that you will be familiar with e.g. power off stalls. This project is being undertaken as part of the requirements for Lana Kinney's Masters of Science. This study will help improve the understanding of pilot behavioural reactions.

**What Types of Participants are being sought?**

We are seeking at least 20 local general aviation pilots. Pilots can be at any level of training, however they must have solo flight ability. All pilots will be offered a monetary reimbursement for participation, and can indicate to the researcher whether they would like a copy of the experimental findings when the project is finished.

**What will Participants be asked to do?**

Should you agree to take part in this project, you will be asked to fly eight short flights around New Zealand in the psychology department's flight simulator. The experiment will take around 1 hour and 45minutes. You will be given the aeronautical maps, and flight information before each flight, and then will be asked to complete each flight. During each flight heart rate will be recorded with an unobtrusive ear clip, eye movements will be recorded, and so will the flight performance data.
Please be aware that you may decide not to take part in the project without any disadvantage to yourself.

**What Data or Information will be collected and what use will be made of it?**

Your eye movements, heart rate, and flight data during the flights will be collected and stored electronically. The questionnaire will ask for your name, age, and flight experience information. The data will be used to examine your responses to different flight conditions. Apart from your name, and age, no other personal information will be collected. The researchers below will have access to the data. Results from this project may be published and will be available in the University of Otago Library (Dunedin, New Zealand), but every attempt will be made to preserve your anonymity. You will not be identifiable in the publications. The data collected will be securely stored in such a way that only those mentioned below will be able to gain access to it. At the end of the project any personal information will be destroyed immediately except that, as required by the University's research policy, any raw data on which the results of the project depend on will be retained in secure storage for at least five years, after which it will be destroyed. You can choose to be provided with a summary of the study's results if you wish.

**Can Participants change their mind and withdraw from the project?**

You may withdraw from participation in the project at any time during the experiment and without any disadvantage to yourself.

**What if Participants have any Questions?**

If you have any questions about our project, either now or in the future, please feel free to contact either:-

*Lana Kinney*
Department of Psychology
Telephone Number: 027 399 7122
Email Address: lanakinney92@gmail.com

or

*David O'Hare*
Department of Psychology
Telephone Number: (64)3-479-7643
Email Address: ohare@psy.otago.ac.nz

This study has been approved by the Department stated above. However, if you have any concerns about the ethical conduct of the research you may contact the University of Otago Human Ethics Committee through the Human Ethics Committee Administrator (ph 03 479-8256). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome.

Appendix F. Consent Form for Pilot Participants

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

***PILOT REACTIONS TO DIFFERIENTIAL FLIGHT CONDITIONS.***
***CONSENT FORM FOR PARTICIPANTS***

I have read the Information Sheet concerning this project and understand what it is about. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage.
I know that:-

6. My participation in the project is entirely voluntary;

7. I am free to withdraw from the project at any time without any disadvantage;

8. Personal identifying information (name and age) will be destroyed at the conclusion of the project but any raw data on which the results of the project depend will be retained in secure storage for at least five years.

9. There is no foreseeable discomfort or distress for the participant while taking part in this study.

10. Pilots will received $40.00 for their time and effort. As well as this pilots will gain one entry into a draw for one hour of flight time at a nearby Aero Club. This draw will consist of only the pilots that participate in the study.

11. The results of the project may be published and will be available in the University of Otago Library (Dunedin, New Zealand) but every attempt will be made to preserve my anonymity.

I agree to take part in this project.

............................................................................ ..............................
(Signature of participant) (Date)

............................................................................
(Printed Name)

Appendix G. Survey Monkey Questionnaire for Pilot Participants

# Pilot Information Survey

Pilot Information

* 1 What is your name?

* 2 What is your age?

* 3 What is the highest level of education you have completed?

- Secondary/High School NOT completed
- Secondary/High School completed
- Bachelors Degree or equivalent
- PhD/Masters/Postgraduate

* 4 Total Flying Hours:

* 5 Years of flying experience:

* 6 Approximate hours of flying in the last 12 months:

__________

* 7 Current licences held: (please tick highest level)

- ◯ Airline transport
- ◯ Commercial
- ◯ Private
- ◯ Student

* 8 Current Ratings held: (Please tick all that apply)

- ☐ Instrument
- ☐ Multi-engine
- ☐ Helicopter
- ☐ Flight Instructor

Other (please specify)

__________

* 9 In which category did you undertake the majority of your flight hours over the last 12 months?

- ◯ Recreational
- ◯ Light commuter/charter
- ◯ Corporate
- ◯ Airline
- ◯ Military
- ◯ Flight Instruction
- ◯ Aerial work/Agricultural
- ◯ Emergency Services
- ◯ Test flying

Other (please specify)

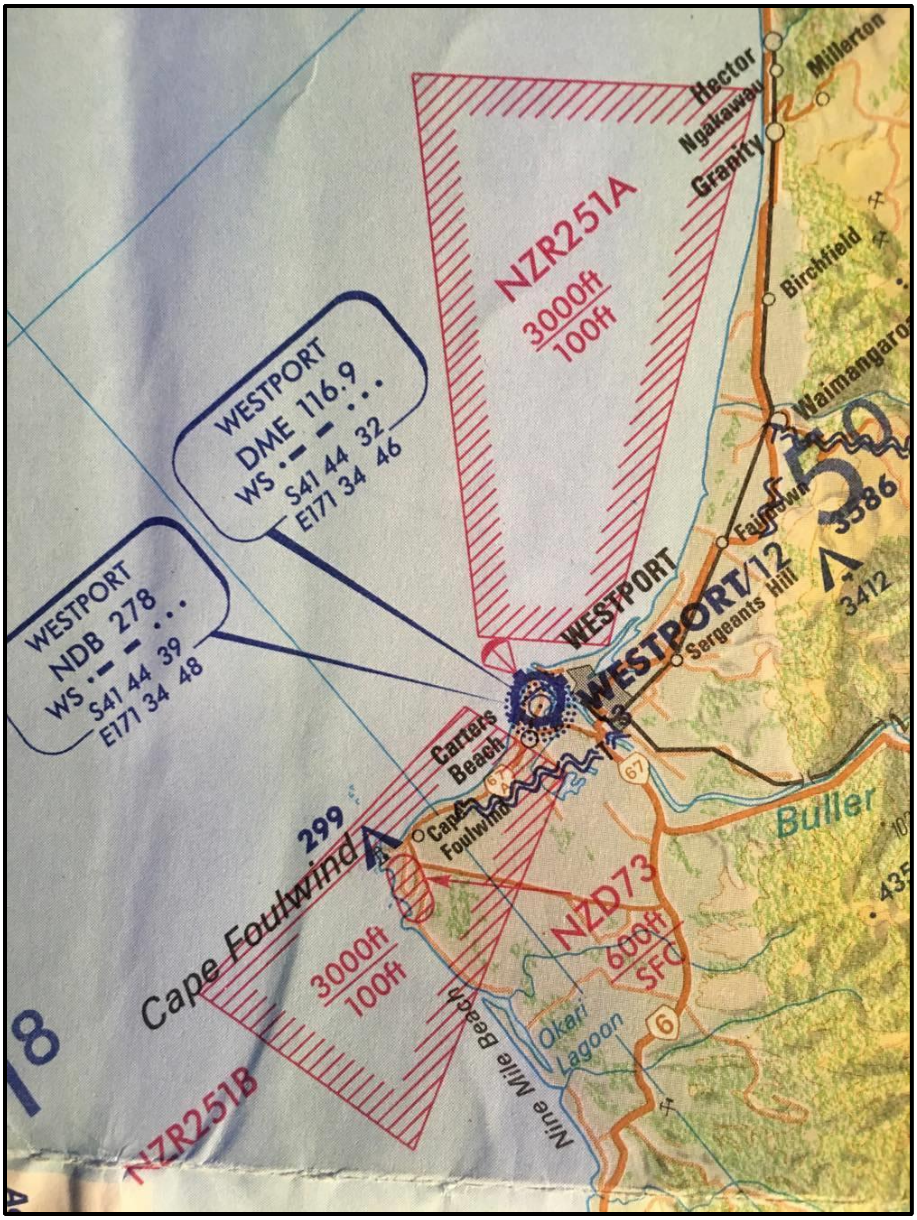
10 Do you take any medications that may affect your heart rate?

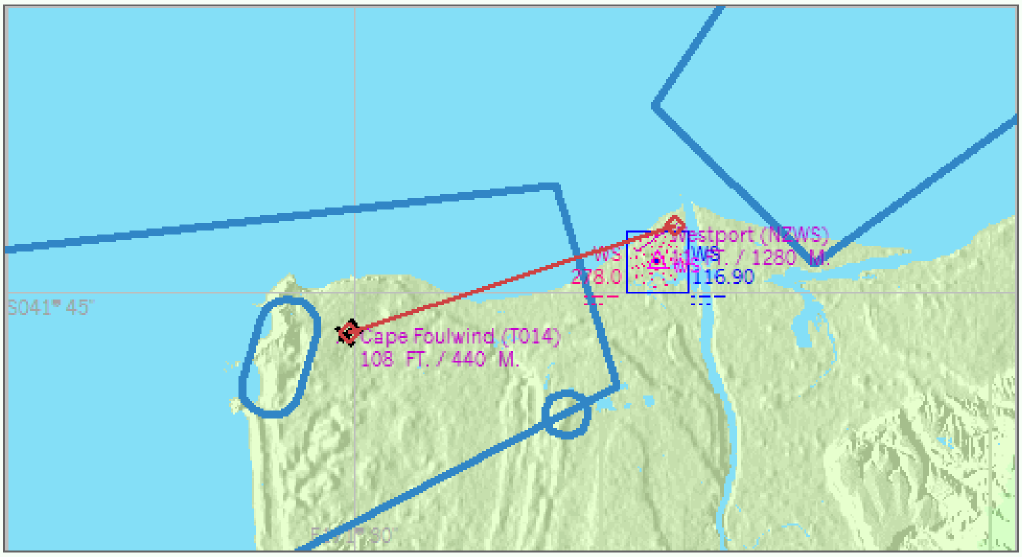
- ◯ No
- ◯ Yes

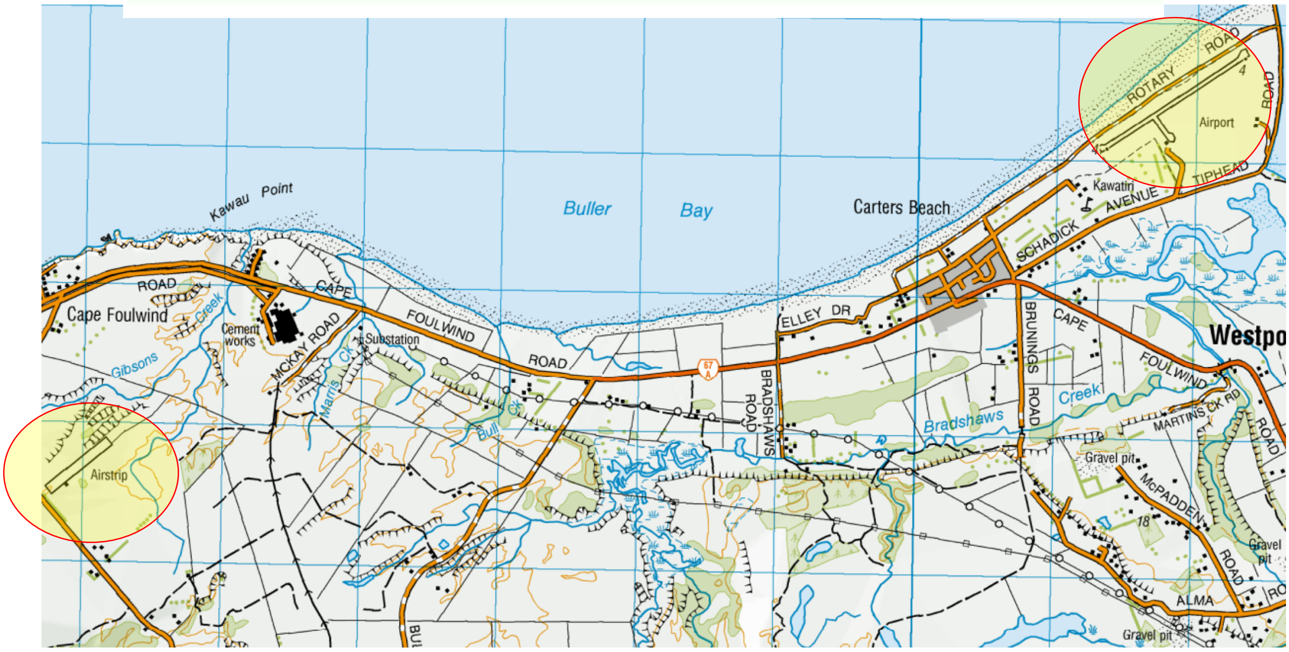
Appendix H. Flight Information Sheet for Pilot Participants

## Cape Foulwind to Westport

Your first flight will be a short flight from Cape Foulwind to Westport. Your job is to simply move the aircraft from the airstrip at Cape Foulwind to the adjacent Westport airstrip. Cape Foulwind is a prominent headland on the West Coast of the South Island, overlooking the Tasman Sea. It is located ten kilometres west of the town of Westport. Make sure when you land you join downwind for 22 (see Aerodrome chart).






**Flight information:**

Distance: 8.6km

Estimated fuel burn: 1.61/1.1kg

Estimated time en route: 0.02

Direct Heading: 049

**Weather Forecast**

Wind 162° at 5kts

Visibility 80km

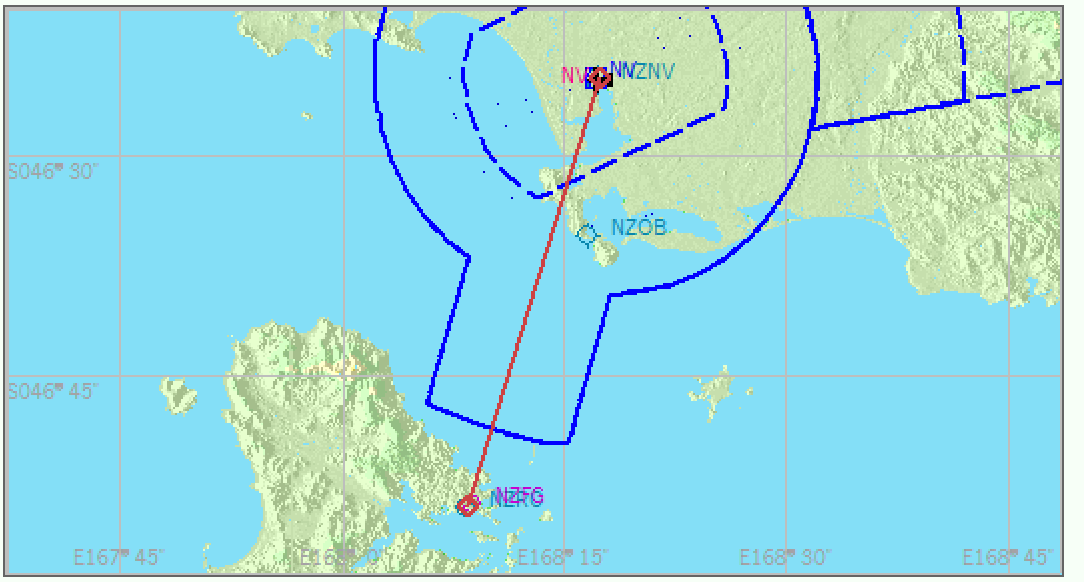
Light clouds

Temperature 15°C, Dew point 10°C

QNH 998

# Invercargill to Ryan’s Creek (Stewart Island)\

You will take off from Invercargill airport, flying directly into a strong southerly which is common at this airport. Your goal is to fly the aircraft to Ryan’s Creek aerodrome situated on Stewart Island. The position of your aircraft is indicated on the aerodrome chart, you will need to taxi out to the runway where you can then complete an intersection take-off.




**Flight information:**

Distance: 56.1km

Estimated fuel burn: 14.41/10.3kg

Estimated time en route: 0.21

Direct Heading: 171

**Weather Forecast**

Wind 218° at 40kts

Visibility 64km

Light clouds

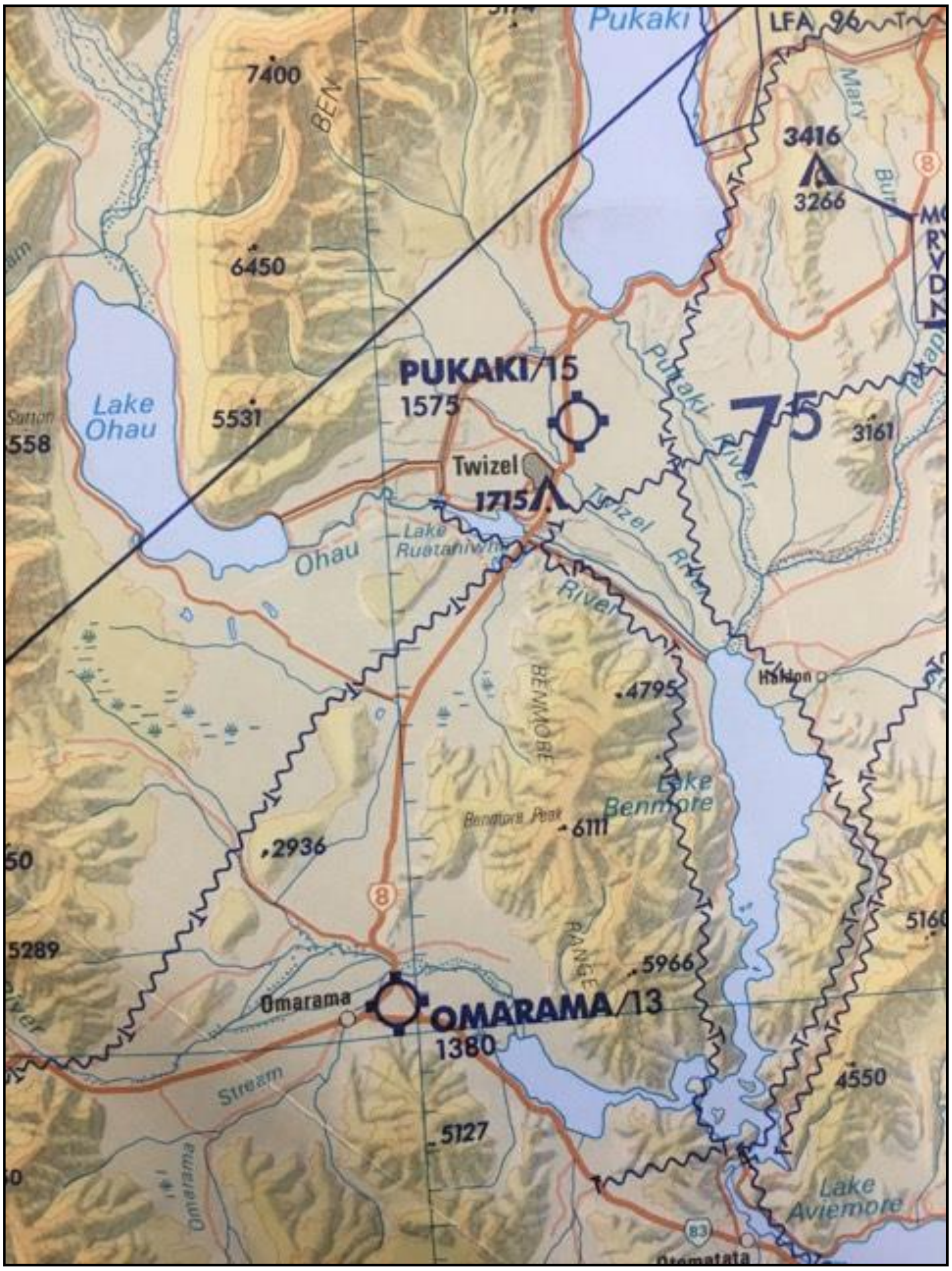
Temperature 7°C, Dew point 4°C

QNH 1013

**Warning: Due to strong winds make sure you taxi out to the runway carefully.**

# Pukaki to Omarama

In this flight you are instructed to take off from Pukaki airport and land at Omarama airport. However, at some point into the flight there will be a programmed engine failure which will mean that you will have to complete a forced landing. The aircraft is positioned at the southern end of the runway 33.




**Flight information:**

Distance: 29.1km

Estimated fuel burn: 5.41/3.9 kg

Estimated time en route: 0.08

Direct Heading: 176

**Weather Forecast**

Wind 334° at 5kts gusting to 8kts

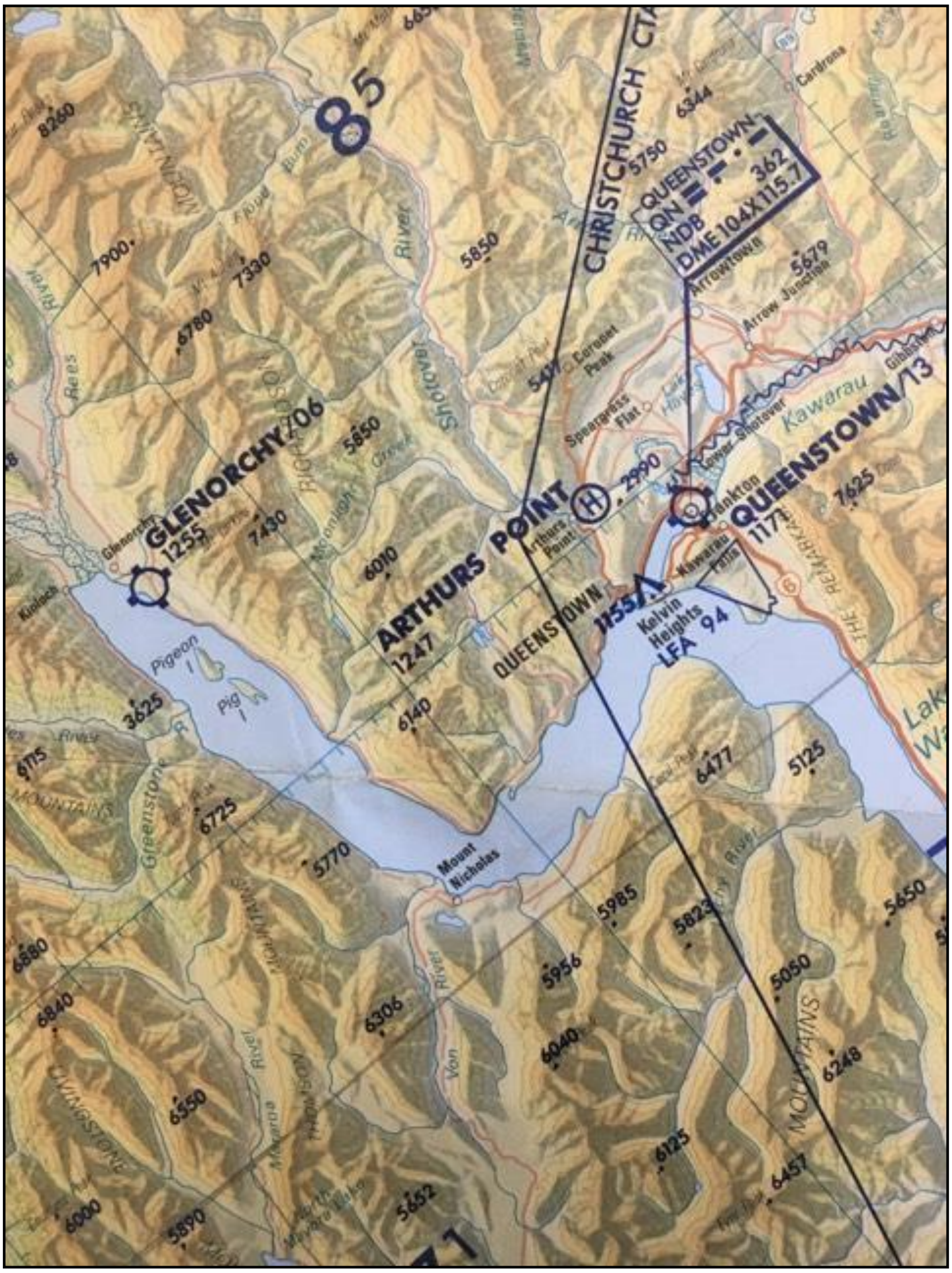
Light Cloud

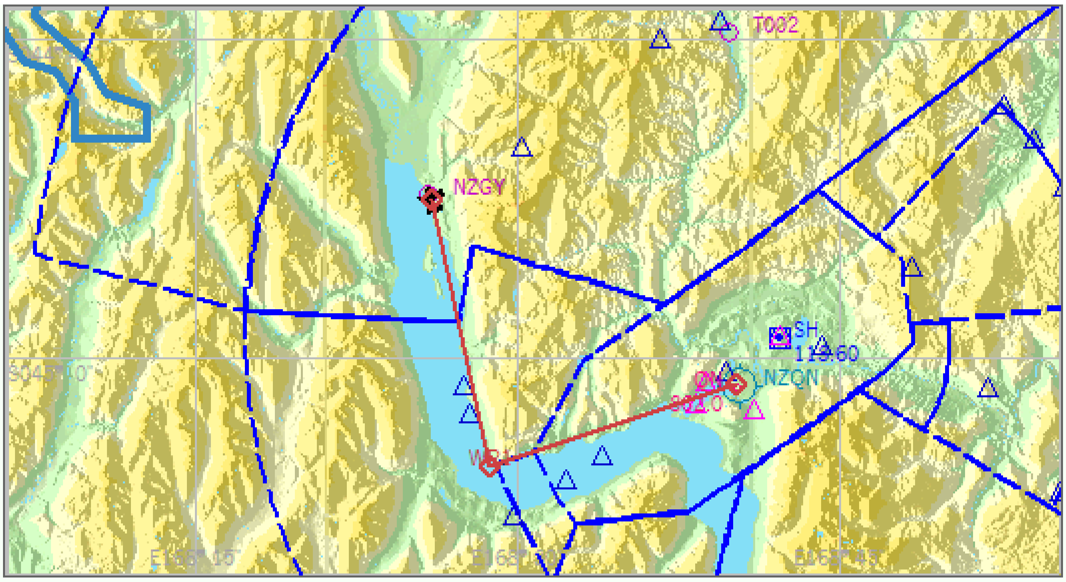
Temperature 21°C, Dew point 4°C

QNH 1013

## Glenorchy to Queenstown

This flight is a scenic passenger flight between Glenorchy and Queenstown. You should fly the aircraft as if you were a paid tour guide. You will take off from Glenorchy airstrip and follow along Lake Wakatipu, and finally land at Queenstown airport. The aircraft is positioned at the southern end of runway 32.






**Flight information:**

Distance: 46.2km

Estimated fuel burn: 8.01/5.8 kg

Estimated time en route: 0.12

Direct Heading:

- Waypoint 1: 143
- NZQN (airport): 047

**Weather Forecast**

Wind 317° at 5kts

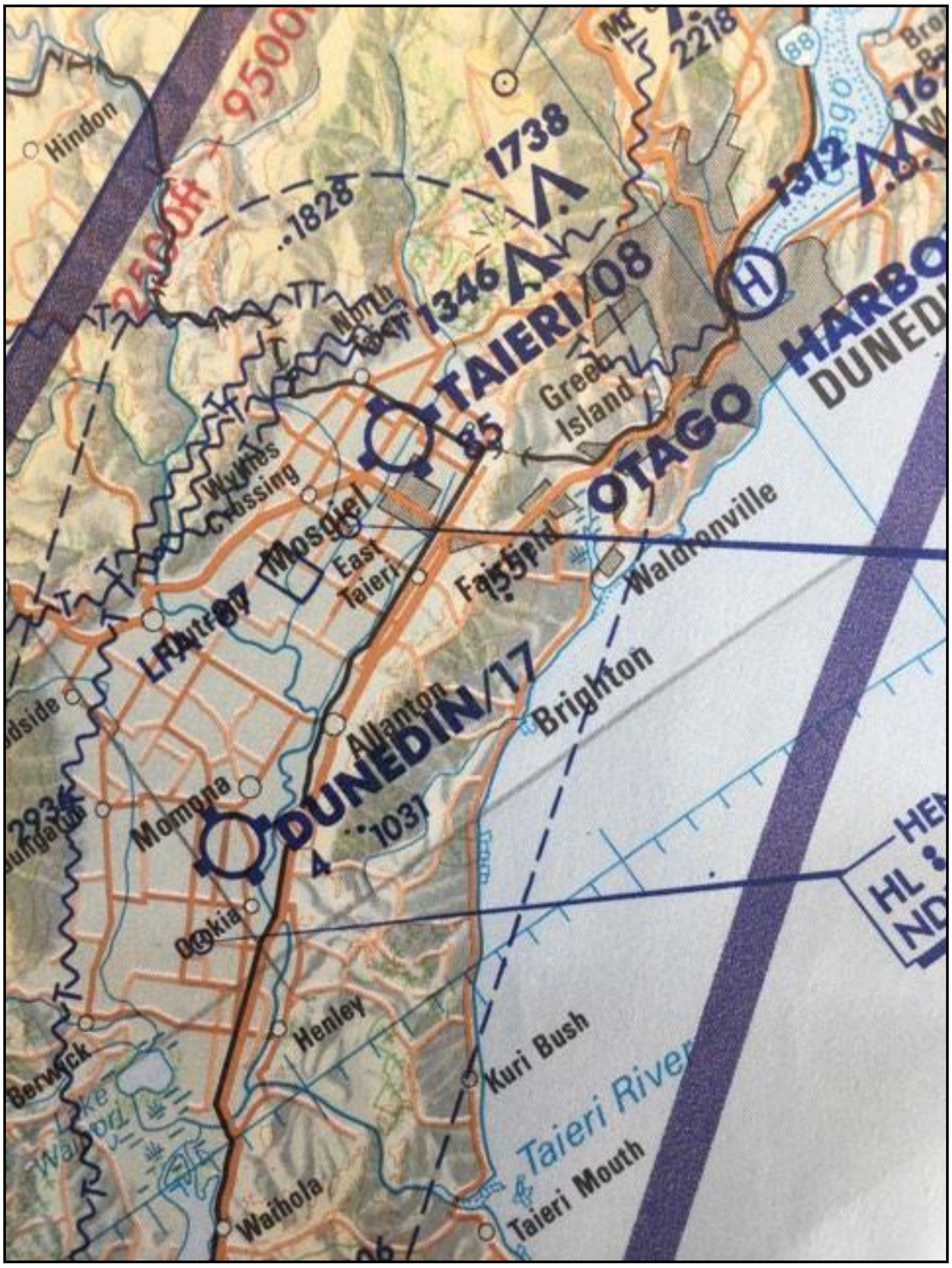
Light Clouds

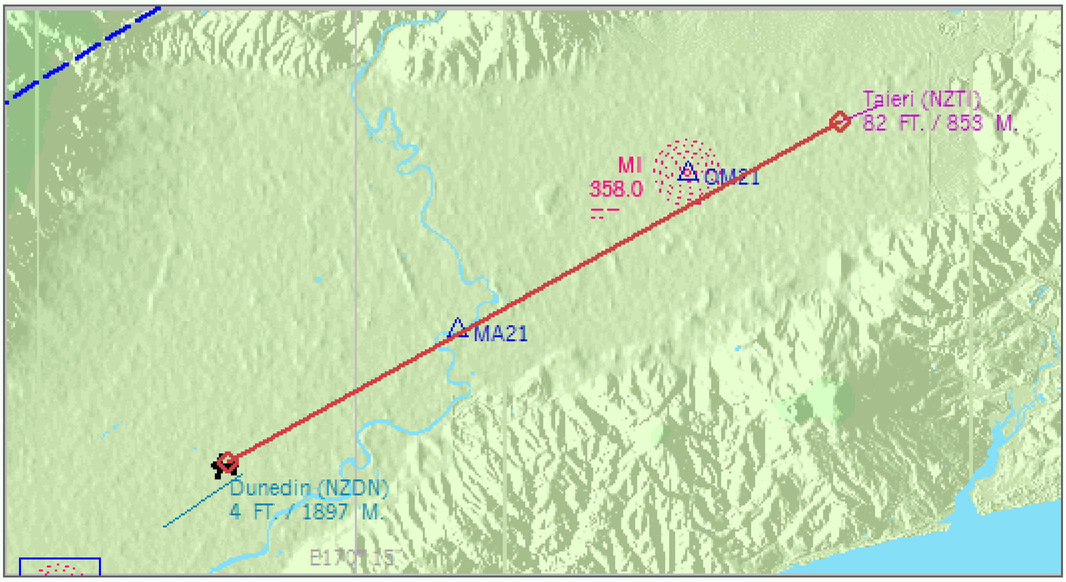
Temperature 15°C, Dew point 5°C

QNH 1013

## Dunedin to Taieri

This is a flight in which you will take off from Dunedin airport, you will fly to an area and altitude where you feel safe to then practice an engine-off aerodynamic stall. Practice this stall just as you would in flight training. Once you have completed the stall and recovered you should proceed and land at Taieri airport. The aircraft is parked outside Mainland air hangar. You will need to taxi to the runway and take-off.






**Flight information:**

Distance: 13.6km

Estimated fuel burn: 2.51/1.8kg

Estimated time en route: 0.03

Direct Heading: 035

**Weather Forecast**

Wind 094° at 2 kts gusting to 6 kts

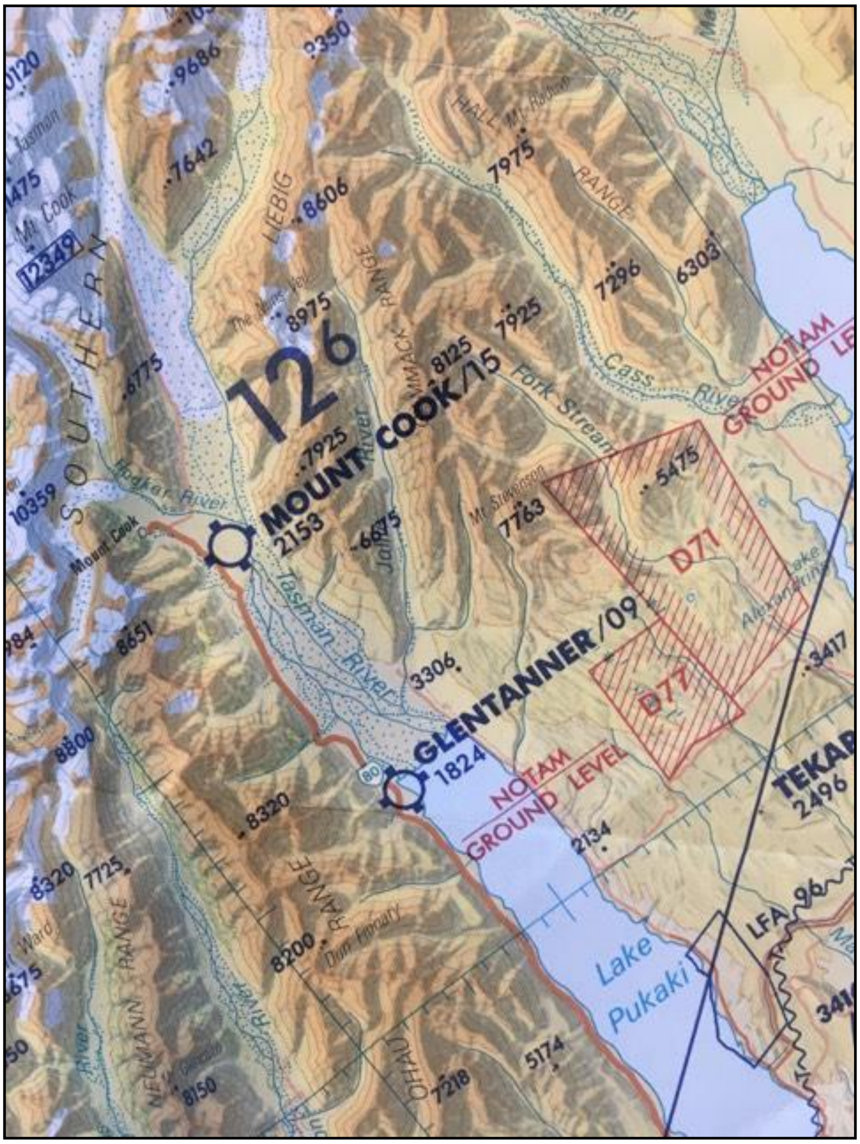
Scattered clouds

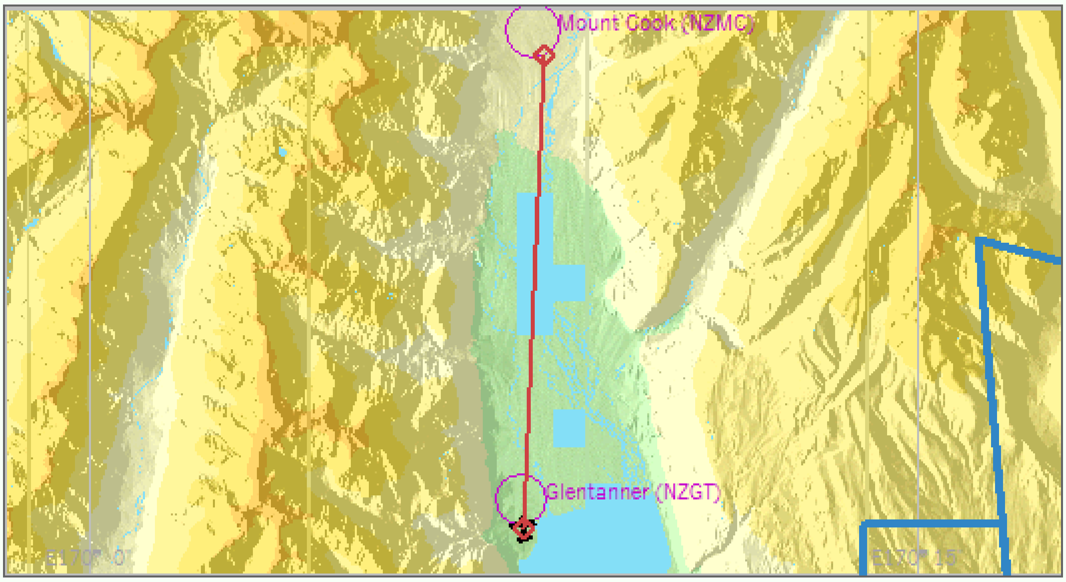
Temperature 19°C, Dew point 4°C

QNH 1013

## Glentanner to Mount Cook

This is a short flight from Glentanner airstrip to a neighbouring airstrip at Mount Cook. Your goal is to move the aircraft to where it is needed at Mount Cook. The aircraft is positioned at the Southern end of runway 33 at Glentanner. Position yourself for a straight-in approach to 31 at Mount Cook.






**Flight information:**

Distance: 15.9km

Estimated fuel burn: .01/2.1kg

Estimated time en route: 0.04

Direct Heading: 339

**Weather Forecast**

Wind Calm

Visibility 16km

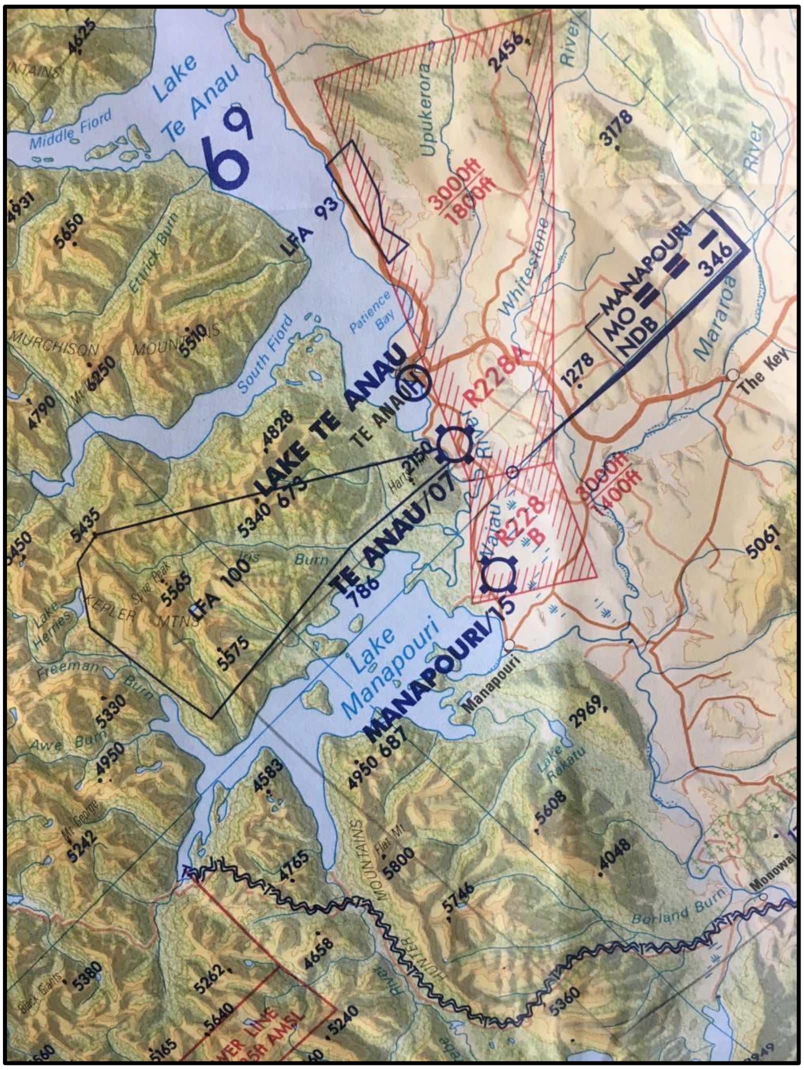
Light clouds

Temperature 2°C, Dew Point 10°C

QNH 1013

## Manapouri to Te Anau

This is a short positioning flight from Manapouri to a neighbouring airstrip at Te Anau. The aircraft is positioned at the eastern end of runway 26 at Manapouri.





**Flight information:**

Distance: 8.5km

Estimated fuel burn: 1.61/1.2 kg

Estimated time en route: 0.02

Direct Heading: 356

**Weather Forecast**

Wind 319$^{o}$ at 10kts gusting to 16kts

Visibility 64km

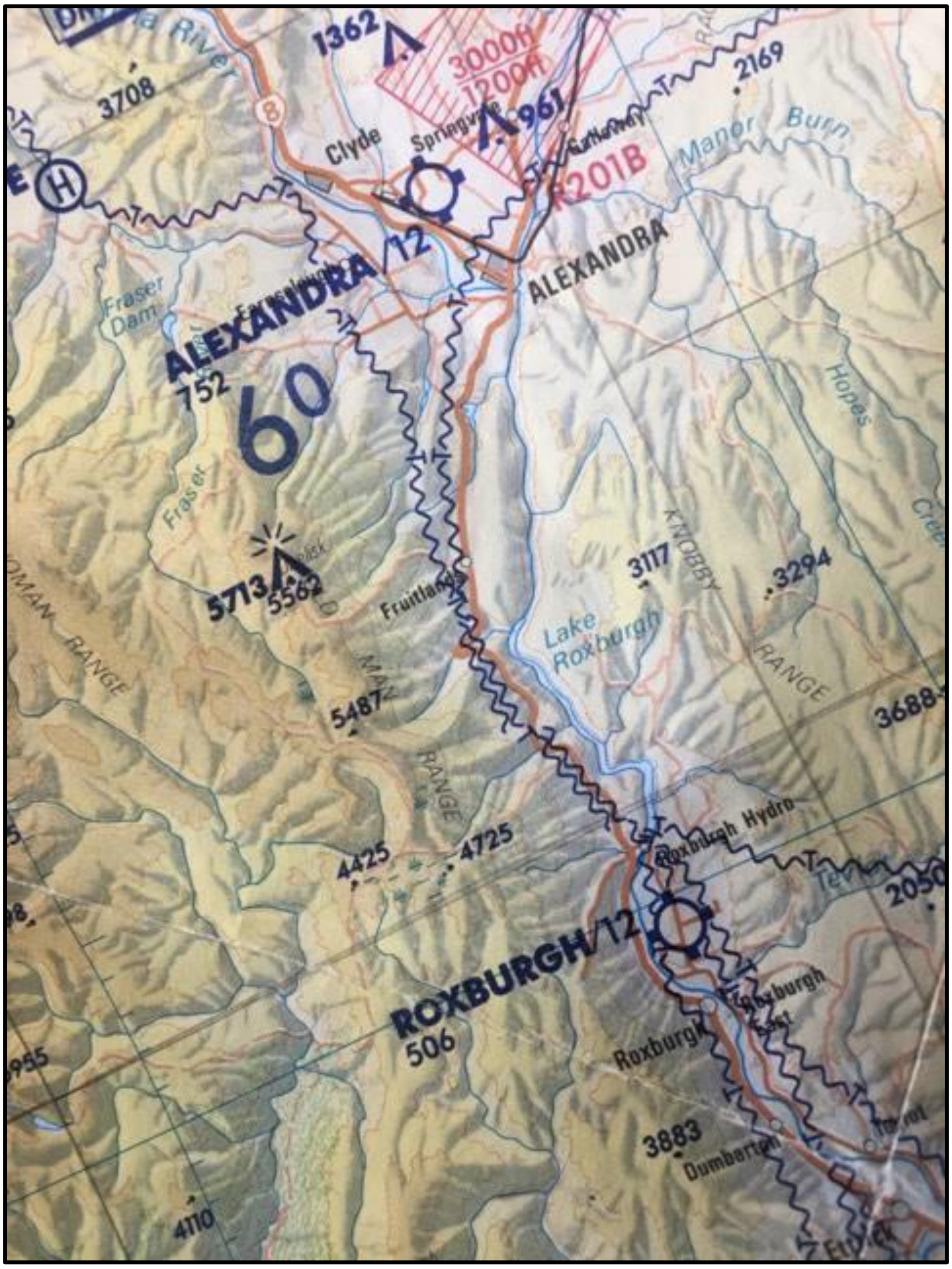
Light Clouds

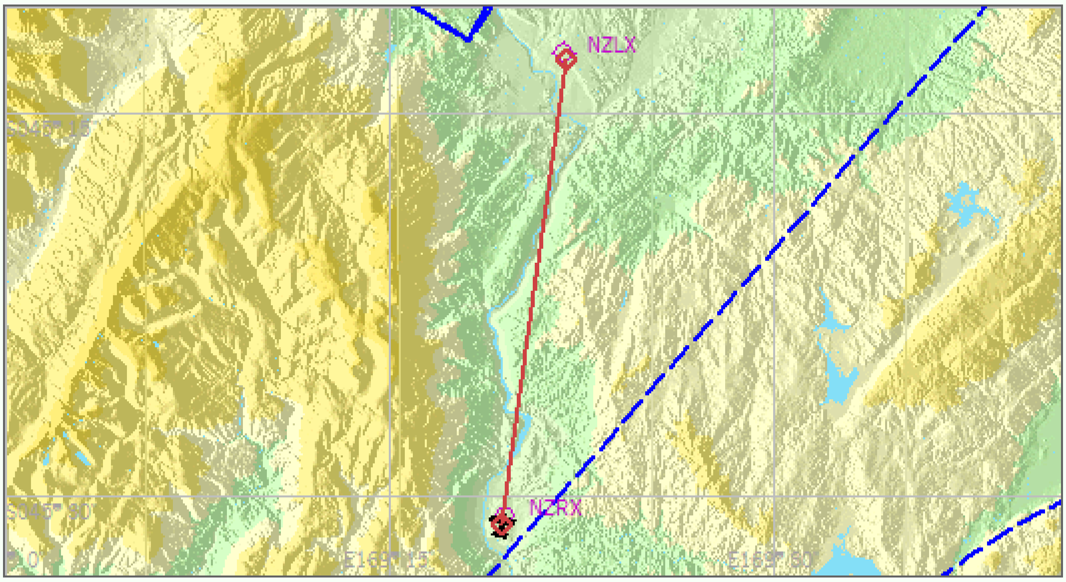
Temperature 15$^{o}$C, Dew point 5$^{o}$C

QNH 1013

## Roxburgh to Alexandra

This is a short flight in central Otago where you will fly from a rural airstrip in Roxburgh to Alexandra. The aircraft is positioned at the southern end of runway 34 at Roxburgh. When landing at Alexandra, positon yourself for a straight in approach to 32.






**Flight information:**

Distance: 33.8km

Estimated fuel burn: 6.91/5.0kg

Estimated time en route: 0.10

Direct Heading: 343

**Weather Forecast**

Wind 336° at 10kts gusting to 20 kts

Visibility: 80km

Broken clouds

Temperature 6°C, Dew point -3°C

QNH 1027

# Appendix I. AOI Collection Layout example


Inside
G-BAFM