Digital vs Paper-Based Checklists in High Performance Single Pilot Aircraft: A Mixed Methods Investigation.

A 190.895 (60 credit) research report presented in partial fulfilment of the requirements for the degree of

Master of Aviation

At Massey University, Palmerston North, New Zealand

Peter Barron

October 2022

Abstract

The aircraft checklist has been described as the most critical man-machine interface in aviation. Checklists can significantly enhance flight safety when designed well and used properly. The 'look' and content of a checklist affects how well pilots interact with it. Emergency checklists in particular, are only accessed during emergency situations, a time of heightened stress and cognitive degradation for the pilot. It is crucial therefore, that emergency checklists are developed with precision and skill to facilitate ease of use during times of stress. A poorly designed checklist can hinder rectification of an emergency situation and can adversely affect flight safety.

The aim of this study was to determine whether flying performance is improved, and pilot workload is lower, when using a digital checklist application created specifically for the T-6C Texan II compared with the existing paper-based Quick Reference Handbook (QRH). For this study, twenty pilots from the Royal New Zealand Air Force (RNZAF) underwent two emergency scenarios in a flight simulator using either the QRH or the digital checklist application. The independent variable was checklist type (paper or digital). Dependent variables were: i) time to find the checklist; ii) time to complete the checklist; iii) flight path accuracy; and iv) workload. Additionally, a qualitative investigation into error occurrences during checklist execution was undertaken.

The results suggested that workload is lower when using the digital checklist application compared to the QRH, but there were mixed results regarding the improvement in flying performance with the digital checklist application. The time to find the checklist was quicker with the digital application, but checklist completion times and flight path accuracy were similar across both checklist types. The qualitative investigation noted that the digital checklist reduced errors and was easier to manipulate. The collection of qualitative data enabled the generation of a hypothesis that frequency and type of error occurrences are affected by checklist type.

Despite the interface improvements of the digital checklist over the QRH, this research suggests that an enhanced checklist interface is secondary to checklist location or checklist content, and that the greatest gains in safety will likely be achieved by addressing these two factors over checklist interface.

This research provides support for an iPad mount in front of the pilot in the T-6C. Additionally, this research provides further evidence that the T-6C checklist content is poorly written and can negatively impact flight safety and may assist in arguing for a content re-write. From a wider perspective, most RNZAF pilots fly with a kneeboard and this research may be relevant for other aircraft types operated

by the RNZAF, noting also that other aircraft types are flown by two pilots which may negate some of the findings in this research.

Further research should standardise the placement of the checklist to fully determine the relationship between flying performance and checklist type. Additionally, future research could also make use of eye tracking equipment to measure attention switching and could investigate the hypothesis generated from the qualitative data.

Acknowledgements

I would like to thank the staff and students of No. 14 Squadron RNZAF for both providing me the opportunity to conduct this study and for participating in it, and for making available one of the two highly utilised flight simulators for duration of the experiment.

This research would not have been possible without another instructor, Squadron Leader James Peters, to act as a second researcher and to help facilitate the simulator sessions. Jimmy, I enjoyed our discussions, and the success of this study was largely down to your organisation.

I would also like to thank Dr Steve Jarvis, of Jarvis Bagshaw Ltd, UK, for his comments and input during the writing of this report, and for his approval to reference his work.

Finally, I would like to thank my research supervisor, Dr Andrew Gilbey for his guidance, feedback, and support throughout this endeavour.

Table of Contents

Definitions and Abbreviations

Caution An operating procedure, technique, etc., which could result in damage to equipment if not carefully followed.

EEC Electronic Emergency Checklist. A digital copy of the T-6C QRH, with interactive functionality.

Flight Director A flight director is a guidance aid that is overlaid on the attitude indicator and shows the pilot the attitude required to follow a certain trajectory.

ICAO International Civil Aviation Organisation.

Note An operating procedure, technique, etc., which is considered essential to emphasise.

QRH Quick Reference Handbook. A paper-based document which contains procedures applicable for emergency situations.

Trim To 'trim' an aircraft is to adjust the aerodynamic forces on the control surfaces so that the aircraft maintains the set attitude without any control input.

Upset Attitude An undesired aircraft state characterised by unintentional divergences from desired flight path parameters, which may involve pitch and/or bank angle divergences, or airspeed divergences.

Warning An operating procedure, technique, etc., which could result in personal injury or loss of life if not carefully followed.

1. Introduction

The aircraft checklist has been described as “the most critical man-machine interface in aviation” (Doherty, 2001, p. 30). When checklists are designed well and used properly, they significantly enhance flight safety (Ross, 2004). In the last three decades in particular, increasing attention has been given to checklist design by the human factors community, and in fact the match between the human and the aviation system is now considered so crucial that human factors principles are integrated throughout the aircraft design process (Bridger, 2017; Wiegmann & Shappell, 2003).

It is important that checklists are tools which help rather than hinder the pilot. Careful consideration must be given to their design and content, otherwise their usefulness as a safety tool is lost. Moreover, poorly designed checklists can create safety problems themselves, and suddenly a tool which was intended to enhance safety now actively detracts from it. Despite the increasing human factors attention given to checklists, often this principle is still not fully considered by aircraft designers (Burian, 2006). A human factors investigation undertaken by Jarvis Bagshaw Ltd, UK, in 2017 argued that the checklist that is the subject of this research report is no exception.

The underlying theme of this research report is that as pilot workload is increased, due to some sort of aircraft fault necessitating the use of the emergency checklist in flight, flying performance decreases. In situations such as these, attention is no longer solely required to fly the aircraft, but also to manipulate and read the checklist, and manage the emergency (Wickens, 2017). A checklist that places undue demands on a pilot’s mental resources during an emergency situation, especially in a single pilot, manually flown aircraft, diverts attention away from aircraft control and substantially increases the likelihood of an unsafe situation developing (Jarvis, 2017). A checklist which requires lower pilot workload to operate should therefore result in a performance increase.

The Royal New Zealand Air Force (RNZAF) operates a fleet of T-6C Texan II aircraft as its pilot training platform. The aircraft has no autopilot and is often flown with a single pilot. It has been suggested that the manufacturer provided paper checklist is poorly written and unwieldy, and in times of an emergency it may hinder the pilot in achieving a safe outcome (Jarvis, 2017). An electronic checklist was developed by the author to address the perceived shortcomings of the paper checklist. The research in this report used a mixed methods approach to determine whether the electronic checklist reduces pilot workload and improves flying performance when compared to the paper-based checklist.

The literature review that follows begins with a discussion on checklists, before moving onto checklist design principles. Following this, the concept of human factors is briefly covered to explain how and where the checklist fits into the human-machine interface. The literature review concludes with a discussion on workload measurement.

Lastly, throughout this investigation the reader should keep in mind the following quote from Lysaght et al. (1989), written for the US Army Research Institute, as it summarises the underlying reason for this research, particularly in the context of military operations:

> Performance is what we are ultimately concerned with, can the operator successfully complete the mission?... Not only do we not want the mission to fail, we also do not want the man or machine to be damaged. Having anticipated and predicted a trouble spot, the… goal is to correct those situations in which performance falls (p. 9).

2. Literature Review

2.1 Checklists in Aviation

Problems with checklists and checklist use have a long history in aviation accidents (Dismukes & Berman, 2008). In 1990, after decades of relative neglect in the human factors arena, Degani & Wiener (1990) conducted one of the first in-detail human factors analyses of aviation checklists. With the benefit of hindsight, and given that the aviation industry has been formally using checklists since 1935[1] following the crash of Boeing's Model 299 (what would become the B-17 Flying Fortress), it is surprising that the subject of checklists escaped the scrutiny of the human factors profession for so long.

Degani & Wiener's (1990) investigation resulted in a raft of recommendations for checklist design, operating philosophies, and pilot/checklist interaction, and concluded that, "the unique interaction between checklists, humans, machines, and the operational environment, makes the checklist problem a true human factors issue" (p. 60).

So, what are checklists? Checklists - while a simple concept - have been variously defined. The Federal Aviation Administration[2] defined a checklist as "a formal list used to identify, schedule, compare, or verify a group of elements or actions". Hales & Pronovost (2006) described a checklist as "a list of action items or criteria arranged in a systematic manner, allowing the user to record the presence/absence of the individual items listed to ensure that all are considered or completed" (p. 231). Definitions of checklists abound; however, the central idea is that a checklist consists of a series of ordered steps which are actioned by the flight crew to achieve safe operations.

Checklists are an important tool in error management, with the aviation industry coming to rely heavily on them due to their effectiveness in enabling significant reductions in human error and improvements in operational safety (Hales & Pronovost, 2006). Their safety benefits have been realised as a tool to standardise operating procedures (Orlady & Orlady, 1999), and to ensure that the aircraft is configured appropriately for any given phase of flight (Degani & Wiener, 1993). In short, the

---

[1] For a one page history of aviation checklists, see Appendix E of 'Ludders, J. W., & McMillan, M. (2016). *Errors in veterinary anaesthesia*. John Wiley & Sons'.

[2] FAA Flight Standard Information Management System (FSIMS). Volume 3, Air Operator Technical Administration; Chapter 15: Manuals, Procedures, and Checklists; Section 5: Aircraft Checklists. https://fsims.faa.gov/WDocs/8400.10%20Air%20Transp%20Ops%20Insp%20Handbk/Volume%203.%20AIR%20OPERATOR%20TECHNICAL%20ADMINISTRATION/Vol%203-Chap%2015-Sec%205.htm

introduction and development of the aviation checklist has been a hugely successful safety innovation (Higgins & Boorman, 2016).

Despite their large scale success and influence on safety, however, if pilots are not committed to their use, then they have no safety value. Velázquez et al. (2015) examined 83 accident reports between 1988 and 2006 attributed to flight crew error and categorised them according to the Federal Aviation Administration's 'behavioural traps'. They categorised over 75 percent of accident reports against "neglect of flight planning, pre-flight inspections, and checklists" (p. 4). To mitigate checklist neglect, Degani & Wiener (1990) argued that checklist design should promote a positive attitude from pilots toward their use.

A positive attitude to checklists is especially important with regard to emergency checklists – the subject of this research report. Pilot's attitudes to checklists are directly affected by their content and design, and therefore emergency checklists in particular must be developed with "precision and skill" (Burian et al., 2005, p. 11). While it is hoped that they are never needed, when they are required, they become a crucial tool in emergency management (Hales & Pronovost, 2006).

Traditionally, checklists have tended to be paper-based, compiled in a document called the Quick Reference Handbook (QRH) which contains both normal and emergency[3] checklists. There are several disadvantages to paper checklists, a main one being that they are often difficult to access, and this can pose a distraction to primary flying duties (Doherty, 2001). Other disadvantages are the inability to distinguish between completed and non-completed items, the tendency to lose one's place when dealing with competing demands for attention, especially during emergency situations (Burian et al., 2005), the need to occupy one hand when holding the checklist, and the difficulty in reading at night under low cockpit illumination (Degani & Wiener, 1990).

2.2 Electronic Checklists

The proliferation of portable electronic devices such as Apple's iPad has afforded a natural platform for the migration of paper-based checklists. Electronic checklists remove many of the physical handling problems of a paper-based QRH and are readable under all lighting conditions. Additionally, the interactive features of electronic checklists reduce workload by enabling quick and easy access to the required checklist at the push of a button (Snead et al., 2004). It is argued that electronic checklists have three main benefits over their paper-based cousins: electronic checklists improve information

[3] There exists in aviation both 'emergency' and 'abnormal' checklists. Emergency situations refer to risks to safety of flight or personnel. Abnormal situations are circumstances where it is no longer possible to continue the flight using normal procedures, but no danger exists. The T-6C checklist is divided into 'normal procedures' and 'emergency procedures', and in keeping with the manufacturer's convention the term emergency checklist will be used throughout this report.

processing capacity and reduce mental workload; reduce errors; and improve response time (Myers III, 2016).

One of the first electronic checklists to be developed was that of the Boeing 777. The Boeing 777 electronic checklist (ECL), introduced in 1996, has resulted in a 46 percent reduction of errors on the Boeing 777 compared to its paper-based QRH (Boorman, 2000, as cited in Hales & Pronovost, 2006). Boorman (2001) wrote of the Boeing 777 ECL that, “in five years of service experience and over six years of training experience the tool has consistently been found to prevent errors” (p. 1). While the Boeing 777 ECL is built into the flight deck, and not portable like the checklist that is the subject of this study, the comparison is relevant as electronic checklists in general mitigate errors common to paper checklists such as loss of place and item omission[4].

The evolution, in particular of emergency checklists from paper to electronic, has also come with disadvantages. A significant disadvantage is that often digital emergency checklists simply regurgitate the content of the paper-based QRHs, with any inherent limitations or failures in the procedures remaining present (Carim et al., 2022). Additionally, there is still the need for a paper backup due to the possibility of equipment failure (Doherty, 2001). Nevertheless, electronic checklists represent a technological advance and a potential step forward in safety. Much like paper checklists, however, their effectiveness as a safety tool, especially to engender a positive attitude from the pilots that use them, is determined by the quality and attention given to their design.

2.3 Emergency Checklist Design

Emergency situations impose time pressure, high workload, and stress on a pilot, the combination of which can negatively affect cognitive processes (Burian et al., 2003). It is important, therefore, that checklists are complete, clear, easy to use, and efficiently facilitate the pilot in rectifying an emergency scenario (Burian, 2004). Burian (2006) identified three overarching design factors for emergency checklists. These are internal factors, external factors, and how effective the checklist is in aiding the pilot in response to a situation (or, as Bailey Fausset et al. [2020] have termed this last factor, “overall purpose” [p. 140]) .

2.3.1 Internal Factors

Burian (2006) identified 14 interrelated aspects to internal factors. For the purposes of this research paper these items have been condensed into four categories (Table 1): physical properties and

[4] For a detailed summary of paper checklist error modes and how electronic checklists can mitigate them, see Boorman (2001).

usability; layout, format, and display; length, content, and workload; and readability. These categories are each discussed below.

Table 1: Burian's (2006) 14 aspects to internal factors and the author's condensed categories.

Aspects to Internal Factors	**Condensed Category**
Physical properties and interface Organisation and access Navigation, progression, and jumping	Physical properties and usability
Layout, format, and display	Layout, format, and display
Length and workload Engineering completeness Engineering coherence Logical coherence	Length, content, and workload
Typography and use of symbology Nomenclature and abbreviations Language, grammar, and wording Purpose Objectives Level of detail	Readability

2.3.1.1 Physical Properties and Usability

Physical properties include things such as the size and weight of the QRH, the type of materials used and the ability for the checklist to be manipulated (i.e., held in one hand). Usability refers to items such as section dividers, tabs, and other features, and is concerned with how well the checklist is organised, how fast pilots can find the desired checklist, the number of checklists, title of checklists, table of contents, effectiveness of the tabbing system, etc. Overall, usability aspects relate to the ease with which pilots can 'operate' the QRH (Burian, 2006).

2.3.1.2 Layout, format, and display

The 'look' of the checklist and the arrangement of items on a page influence checklist effectiveness. Checklist layout, format and display are critical factors in determining a checklists readability and useability. This category includes aspects such as the number of checklist items per page (Civil Aviation Authority, United Kingdom, 2000), the display and orientation of lists (Wagner et al., 1996), and text justification. On text justification, for example, it is recommended that checklist text should be left justified with a ragged right hand edge, since fully justified text can be hard to read as the spaces between words are uneven (Civil Aviation Authority, United Kingdom, 2000). Burian (2006) notes that often not enough attention is given to this category by checklist designers, resulting in checklists that are hard to read and difficult to follow.

2.3.1.3 Length, content, and workload

Checklist length relates both to the physical length of the checklist and how long it takes to read and complete the associated actions. Degani & Wiener (1993) wrote that, “a long and detailed checklist is no guarantee of safety. Indeed, it carries the risk that some pilots might choose not to use the checklist or may conduct the procedure poorly because of its length” (p. 350). Checklist length must also be adequate to the time available. A comprehensive checklist could be written for any situation, but if the act of completing the checklist takes longer than the time the pilot has available to rectify the situation, then the checklist is pointless (Carim et al., 2022).

On checklist content, a fundamental consideration is deciding which items to include (Degani & Wiener, 1993). Checklist content should not be ambiguous or contain excess verbiage (Turner & Huntly, 1991), since at times of high workload incomplete, ambiguous, or contradictory content is a recipe for pilot error (Burian et al., 2005). An example in the present research is the redundant use of the word ‘inboard’, when referring to landing gear doors. In the aircraft in question, the outboard gear doors are irrelevant to the pilot as they cannot be seen, nor manipulated, and are never actually mentioned in the checklist. Furnishing the pilot with every possible detail can run the risk of distracting the pilot from the core task of flying the aircraft (Jarvis, 2017). Degani & Weiner (1990), in their foundational study, wrote of their surprise at the number of ambiguous content in such a controlled document as the checklist.

Lastly, it is important that checklists are written so as to minimise mental workload, since the time an emergency checklist will be required is also the time that stress levels will be the highest and cognitive performance will be compromised (Johannsen & Rouse, 1983). The Civil Aviation Authority, United Kingdom (2000) provides several suggestions to minimise mental workload such as the avoidance of redundant terms, the need to only present unambiguous information that is relevant to the task, and the avoidance of screen clutter, echoing the “precision and skill” comment by Burian et al. (2005). A more in depth discussion of workload is included under the heading ‘human performance capabilities and limitations under high workload and stress’.

2.3.1.4 Readability

A key aspect of internal factors pertains to the readability of the checklist, determined by font size, font type, use of colours, boldface, italics, and other typographical features. For example, large font is preferred for legibility, while a small font reduces the physical size of the checklist. Degani (1992), in an attempt to provide guidance on checklist design, made recommendations on aspects such as font type and size, the use of lower and upper case lettering, stroke width and height-to-width ratio, horizontal and vertical spacing, and more. Many aviation regulatory bodies now provide recommendations and standards for checklist typography (Civil Aviation Authority, United Kingdom, 2000; European Aviation Safety Authority, 2012; Federal Aviation Administration, 1995). Burian (2004) notes that the subject of typography has probably been the issue most frequently addressed by the human factors community in checklist design.

2.3.2 External Factors

Burian (2006) identified seven aspects to external factors (Table 2). Many of these external aspects are relevant for aircraft manufacturers during checklist design, particularly regarding checklist content. These are collectively categorised here as 'checklist content design' and are outside the scope of this research paper and not discussed. Additionally, while important from an organisational context, social and cultural influences are also not discussed here. The most pertinent and significant factor to this research paper is 'human performance capabilities and limitations under high workload and stress'.

Table 2: Burian's (2006) seven external aspects and the author's condensed category.

Aspects to External Factors	**Condensed categories**
Specific aspects of emergency or abnormal situations Standard operating procedures (SOPs) and aviation regulations Operational requirements Aircraft systems requirements Philosophies and policies, flight crew training, and economic constraints	Checklist content design (manufacturer specific)
Human performance capabilities and limitations under high workload and stress	
Social and cultural influences on crew performance, behaviour, and checklist usage	

2.3.2.1 Human Performance Capabilities and Limitations Under High Workload and Stress

An often underlying cause of aviation accidents is degradation in human performance attributed to the cognitive demands placed on the human (Diehl, 1991). Particularly in emergency situations, high levels of stress can cause cognitive processes to slow, resulting in narrowed attention, impaired decision making, and an overall decrease in performance or operational effectiveness (Kavanagh, 2005), a matter that could be exacerbated by a poorly designed checklist.

The overarching concern in checklist design is to minimise the workload involved in using it. It is, however, hard to limit this discussion to the concept of workload without also discussing attention. Any attempt to explain workload naturally ventures into the concept of attention, since the very act of exercising attention requires the deployment of mental workload (Warm et al., 2017), and the deployment of mental workload is an exercise of attention (Wickens & McCarley, 2019). At the risk of oversimplifying, what follows is a brief summary of the two concepts.

2.3.2.2 Attention and Workload

Theories of attention and workload revolve around the idea that the human has a limited pool of cognitive resources from which to draw upon to achieve a task or tasks. Two main theories exist: The 'Capacity Theory' (Kahneman, 1973), which argues that there is a single general pool of mental resources which can be allocated to various tasks; and the 'Multiple Resource Theory' (Wickens, 1980), which contends that there are several, independent mental resource pools. In multiple resource theory, workload increases if tasks draw from the same resource pool. For example, navigation through unknown terrain interferes with a vehicle control task more so than a verbal communication task, since the vehicle control and navigation tasks draw from the same mental resource pool (Sirevaag et al., 1993). A similar example is flying an aircraft and reading a checklist.

Both approaches contend that as task demands increase, more mental resources are consumed and performance on the task(s) decreases. Inherent in these theories is the requirement for the division of attention amongst tasks. Performance on complex (multiple) tasks, such as operating an aircraft, requires more division of attention than a single simple task, and consequently reductions in performance occur sooner (Epling, 2017). As the limits of human processing capacity are approached it becomes difficult to divide attention (also called task management) effectively (Kantowitz & Casper, 2017), and once task demands exceed available mental resources, overload occurs (Wilkins, 2018), causing lapses in attention, significantly increased error rate, and cognitive lock-up (Wilson & Rajan, 1995).

Wickens (1987) wrote that an "aircraft cockpit is a fascinating environment for the investigation of attention theory precisely because it places such extensive demands on the pilot's capabilities and limitations as a parallel processor of information, the essence of the study of human attention" (p. 608). Tole et al. (1982) highlighted the demands placed on the pilot that Wickens refers to, stating that "the pilot acts as a decision maker, a planner, a manual controller, a monitor, and an event detector" (p. 1). All these items compete for the pilot's attention at once.

Attention can be variously defined as "the amount of resources people devote to a stimuli" (Basil, 1994), p. 180), or, in the case of concurrent tasks, "our ability to do more than one thing at the same time" (Kantowitz & Casper, 2017, p. 169). The second definition echoes Wickens' (1987) description of a pilot as a multi-channel information processor.

Kantowitz (2000) considered mental workload as a subset of attention, and a more useful concept. Mental workload can be defined as "the effort invested by the human operator into task performance" (Hart & Wickens, 1990, p. 258), and "the amount of cognitive or attentional resources being expended at a given point in time" (Charlton, 2019, p. 98). Again, these definitions assume that humans have limited information processing capacity that can be exhausted if there is too much information to process.

The ability to switch attention amongst tasks is affected by mental workload, too. If workload is too high, attention narrows to one task at the expense of other tasks and the ability to switch attention is hindered (Epling, 2017) – the 'cognitive lock-up' mentioned previously. Additionally, the act of attention-switching itself requires mental effort. During complex task performance, for example, Williges & Wierwille, (1979, as cited in Cain, 2007) found that more mental effort is expended managing two tasks simultaneously than the sum of the effort required to perform each task alone.

Research suggests that attention switching is most efficient if it occurs at moments of low workload, when there is an increase in the availability of mental resources (Bailey & Konstan, 2006). Therefore, any checklist designed with the intention of keeping pilot workload low should see a reduction in errors, shorter checklist completion times, and an overall improvement in performance.

The fact that aviation accidents are often the result of poor task management – or poor attention switching (Dismukes & Nowinski, 2007), only serves to reinforce the need for effective human-machine systems in aviation. An important factor in task management on the subject of checklists is the need to minimise 'head down time'.

2.3.2.3 Head Down Time

'Head down time' refers to the inability of a pilot to divide his or her attention optimally between the primary task of flying the aircraft and a secondary task such as reading a checklist. Usually, when the pilot looks away from the primary visual field (the aircraft instruments), they are looking down at maps, charts, checklists, or other documents, hence the term 'head down'. The head down problem is an important safety concern as the longer a pilot's head (or attention) is down, the higher the risk that they may miss critical events such as approaching aircraft or deviations from the desired flight path. Exacerbating the problem is that electronic checklists, particularly those displayed on portable electronic devices, can lead to greater head-down time than paper-based checklists due to the volume of visual information able to be displayed. This is especially problematic in single-pilot aircraft, which are 30% more likely to be involved in an accident than dual-pilot aircraft, as there is no second pilot to pick up monitoring duties (Wilkins, 2018). In the case of the aircraft that is the subject of this research, this problem is made worse since there is no autopilot, and therefore lengthy periods when the pilot's head is down can result in unintentional and undesired flight path deviations.

Head down time can be mitigated however, by ensuring checklists are located in an area that minimises interruption and distraction (Doherty, 2001). This could be on a checklist mount in front of the pilot, for example. A checklist on a portable electronic device lends itself to ergonomic mounting as it requires none of the manipulation that a book (QRH) does, only requires one hand to operate, and provides quick and easy access to the required information.

Good checklist design is also very important in minimising head down time. Burian's (2006) design principles feature here, particularly with regards to the principles of 'layout, format and display' and 'readability'. Doherty (2001), in his study on electronic checklists in fighter aircraft, also notes the importance of displaying and formatting checklist information in a way so as to minimise pilot workload, especially considering that many fighter aircraft are operated with a single pilot.

2.3.2.4 Single Pilot Workload

Much of the literature on pilot workload relates to multi-crew flight decks. In single pilot aircraft however, pilot workload is increased over and above that of multi-crew aircraft. Orasnau & Backer (1996) refer to an experiment conducted by the Hughes Aircraft Company in 1977 where single and two seat aircraft were compared as military pilots flew simulated strike missions. As threats became more numerous, the performance difference between the one and two person crews became greater, with the two person crew consistently showing better performance. In the two crew cockpit, the workload was able to be shared. The experiment's main finding was that, for single pilot aircraft, as workload increases there is a decrease in performance more so than in a dual crew cockpit. In single pilot aircraft especially then, it is important for checklist design to consider the workload it demands from the pilot.

2.3.3 Overall Purpose

The last of Burian's (2006) design factors is that of 'overall purpose'. Overall purpose asks the question, to what degree does the checklist achieve its aim in guiding and directing the pilot(s) in response to an emergency situation? Overall purpose is concerned with enabling effective human performance and is intricately linked to how well internal and external factors have been addressed. As Degani & Wiener (1990) stated, "checklists must be human-centred" (p.60), and the more human centred they are, the greater their effectiveness as a safety tool, and the better they will fulfil their overall purpose.

The previous discussion has centred around checklist design principles and human performance limitations, and while aspects of human factors have been touched on, to better understand where and how a checklist fits from a human factors perspective, a definition of human factors is required.

2.4 Human Factors

"Human Factors is about people. It is about people in their working and living environments. It is about their relationship with machines and equipment, with procedures and with the environment about them" (Hawkins & Orlady, 2017, p. 20). More succinctly, the discipline of Human Factors is concerned with understanding the interactions among humans and other system elements (International Ergonomics Association, 2022). An understanding of interactions enables optimisation of the human into the system and reflects the "human-centred" approach proposed by Degani & Wiener (1990).

The importance of human factors in the aviation industry cannot be overstated. The need for human factors is seen in two broad, interrelated areas. These two areas are the effectiveness (or performance) of the system (which includes safety and efficiency), and the well-being of personnel (ICAO, 1998), and these two areas can be optimally achieved by fitting the system to the human (Dul et al., 2012). So, how exactly do we fit the system to the human?

2.4.1 The SHELL Model

To understand the applicability of human factors in practice, it is useful to use a model. The SHELL model (Figure 1), developed by Edwards in 1972 as the SHEL model, with the second L subsequently added by Hawkins in 1975 (as cited in ICAO, 1998) provides a conceptual model of human factors using blocks to represent the different components of human factors. The human (liveware), at the centre, is the most critical component of the system and interacts with the other system components. The jigsaw-like edges of the system components illustrate that they must be carefully fitted to the human in the centre. Human factors is not concerned with individual blocks per se, but with the interface between the human and the blocks, since it is at the boundaries (or interfaces) of the blocks that problems can occur. Optimising this interface will maximise the safety and performance of the system.

Particularly relevant to this research is the liveware-software interface. This interface is between the human and non-physical aspects of the system such as procedures, manuals, and checklists (ICAO, 1998), and it is this interface that this research paper seeks to optimise.

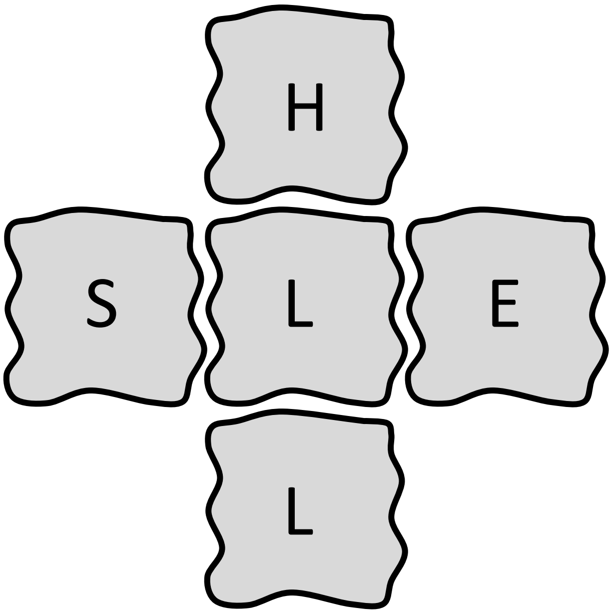


Figure 1: The SHELL Model (Hawkins, 1975, as cited in ICAO, 1998) provides an aid to understanding human factors. The human (liveware) is at the centre, and interacts with the other blocks (software, hardware, and environment, and in some cases with other humans). The diagram does not include interfaces which are do not involve human factors (hardware-hardware; hardware-environment; software-hardware).

Regardless of the quality of the interface, however, the individual who makes up the liveware is subject to considerable variations in performance. The health of the individual, exercise and fitness, fatigue and stress levels, distraction, and many other factors combine to determine the performance effectiveness of the human in the system on a given day, and while human factors is certainly involved with individual well-being, it is particularly concerned with the factors that result in variations in performance (Hawkins & Orlady, 2017), such as attention and workload.

2.5 Measuring Workload

Measuring workload is necessary to quantify the mental effort of performing tasks, and thus the effectiveness of the overall system (Cain, 2007). In the context of an emergency checklist, Burian (2006) argued that a workload evaluation cannot be done in isolation. The workload incurred by operating the checklist can only be gauged by assessing the checklist in the environment for which it was intended – the aircraft. Noting that there are safety implications related to testing new equipment in an aircraft, however, a flight simulator is a good place to start.

Lysaght et al. (1989) noted that workload measurement techniques are usually organised into three broad categories: self-assessment or subjective rating scales; performance measures (primary and secondary); and physiological measures. Subjective rating scales and performance measures have repeatedly demonstrated validity and reliability. Physiological measures have delivered mixed results and are generally not recommended (Cain, 2007; Casner & Gore, 2010; Corwin et al., 1989). Additionally, with the absence of easily available physiological equipment for the author, physiological measures are not discussed in this paper.

2.5.1 Performance Measures

Performance measures assume that performance degrades as workload increases (Kantowitz, 2000). Performance measures of workload can be classified into primary and secondary task measures. Primary task measures attempt to assess the operator's performance on the task of interest directly. Secondary task measures are performed concurrently with primary task measures and help the researcher elicit indirectly how much spare capacity the operator has (Cain, 2007). If the operator can easily perform the secondary task, it can be concluded that the primary task requires only a low or moderate amount of workload. If a breakdown in performance on the primary task occurs while conducting the secondary task, it can be concluded that the primary task places the operator under high or extreme workload, and they are operating at the limit of their mental capacity.

Despite their high validity (Denning, 2003), the main disadvantage for primary measures is that they do not accurately indicate spare mental capacity. For example, a task may be performed equally well by two people, however one person may experience low workload while the other person is overloaded (Sirevaag et al.,1993). If primary performance measures are insufficient, they can be combined with secondary measures.

A disadvantage of secondary task measures is that they must use the same mental resources as the primary task. Therefore, introducing a secondary task may add hidden workload to the primary task, and, for an environment such as aviation, this could have safety implications if conducting tests in an actual aircraft (Sirevaag et al., 1993). Additionally, the feasibility of the secondary task method is limited when the primary task is a real-world activity such as flying since it is difficult to superimpose a secondary task on top of the primary flight task while maintaining an adequate level of realism in the study (Denning, 2003).

2.5.2 Subjective Workload Measures

Subjective measures can be categorised as unidimensional, which provide an overall workload score, and multidimensional, which break workload down into individual variables. Multidimensional scales are more diagnostic, but also more complex and time consuming to complete than unidimensional rating scales. Unidimensional rating scales are quick and easy to administer, minimally distracting, and do not require complicated analysis techniques (Miller, 2001). Furthermore, there is evidence that unidimensional scales are just as accurate as multidimensional scales (Hendy et al., 1993). In addition, it is recommended that structured or unstructured interviews and questionnaires are used to complement unidimensional rating scales to gain a fuller picture of workload (Roscoe, 1987).

There are disadvantages of unidimensional scales. One disadvantage is that individuals experience workload differently for a given task and may consider different variables when deciding on a rating since their personal definitions of workload may vary (Hart & Wickens, 1990). Secondly, since univariate scales produce ordinal data, which cannot be manipulated mathematically, there is some debate when it comes to analysing the results of these scales (Kantowitz & Casper, 2017). The output of unidimensional rating scales provides no context of how much better (or worse) one score is over another, as does interval data provided by multidimensional measures. For example, three sprinters may finish a 100 metre race in first, second and third with respective times of 10.1 seconds, 10.6 seconds, and 10.8 seconds. An ordinal scale would simply rank them as first, second and third and provide no further relationship between their positions, whereas interval data would show that second and third were closer together than first and second. Annett (2002), in an investigation into subjective workload measures, noted that often experimental comparisons require only an ordinal result, and in these cases, simply showing that A is preferred to B is adequate. Alternatively, Cain (2007) argued that subjective workload rating data could be considered as interval data as there is insufficient evidence to the contrary, observing that “the typical approach seems to be to ignore possible violations of mathematical axioms in favour of convenience, accepting the risk that conclusions may not be justified given the data used” (p. 4-7). This research paper will treat data obtained from subjective workload measures as ordinal data.

2.5.3 Workload Measure Summary

For maximum accuracy when measuring workload, it is necessary to use a combination of measures (Miller, 2001; Roscoe, 1987). Measures should include at least one objective measure and make use of quantitative subjective assessments (Cain, 2007). This study assessed workload objectively with performance measures, and subjectively with two univariate rating scales commonly used in the RNZAF for flight test and evaluation. The workload assessment was complemented with unstructured interviews.

2.6 Literature Review Conclusion

Checklists are an effective safety tool in aviation. Traditionally, paper-based checklists have been compiled in a single document called the Quick Reference Handbook, often creating a bulky object which is difficult to manipulate and operate. The proliferation of portable electronic devices has enabled the migration of paper checklists to digital format. Electronic checklists, particularly emergency checklists, mitigate many of the disadvantages of paper checklists. They decrease errors, are easier to use and read, and reduce pilot mental workload. Often, however, checklist content is simply regurgitated into electronic form with any inherent limitations or failures in the procedures remaining present.

Good emergency checklist design can be addressed by considering four (condensed) internal factors: physical properties and useability; layout, format, and display; length, content, and workload; and readability. Emergency checklist design must also address human performance capabilities and limitations under high workload and stress. Poorly worded, ambiguous emergency checklist content adds to pilot workload, and can increase head down time leading to a less safe operating environment. Single pilot operations are inherently more prone to error from high workload than two crew flight decks as workload is unable to be shared.

The checklist and the human operator are two components of an aviation system. 'Human factors' occurs in the boundaries between system components. Optimising the interactions that occur at these boundaries will achieve the twin objectives of the human factors discipline: maximising the effectiveness of the system, and the safety of personnel. These objectives can be best achieved by fitting the system to the human. In the context of the checklist, it must be human centred, and the assessment of workload helps to determine how human-centred and effective checklists are.

Doherty (2001), in his investigation of the efficacy of electronic checklists in the United States Air Force fighter community, wrote that "the design of electronic checklists will not guarantee error-free flying operations. However, it is hoped this application of technology… can help military pilots manage workload better, resulting in lower occurrences of human error within the Air Force community" (p. 34). Exactly the same sentiment applies in the context of this research.

2.7 Research Problem

The T-6C is a high performance, single engine, two seat, training aircraft. The pilots sit in a tandem arrangement, with the rear pilot sitting in a raised cockpit for improved visibility. The T-6C is often flown by a single pilot, however, who must manage all aspects of the flying environment. The aircraft has no autopilot or flight director, is sensitive in pitch and is difficult to trim accurately in this axis (Jarvis, 2017). The elevator trim switch (which controls the pitch trim) can either be pressed momentarily (called a 'blip') or held in position to run the trim motor for longer. The blip provides for a mechanically pre-set amount of trim which, at most airspeeds, is too coarse of an adjustment. The result is that the pilot is often faced with a choice between being trimmed slightly nose up or slightly nose down and is required make continual stick inputs to hold the desired aircraft attitude.

A human factors review of the aircraft in 2017 conducted by Jarvis Bagshaw Ltd, UK, found that the lack of autopilot and undesirable trim characteristics in the T-6C caused an increase in pilot workload by requiring the pilot to constantly switch attention between the primary flying instruments and secondary cockpit tasks such as reading a checklist. This continual attention switching between tasks increased heightened the risk of and pilot error and upset attitudes occurring (Jarvis, 2017).

Emergency situations in flight are dealt with by referring to the aircraft's 'Quick Reference Handbook' (QRH), a hardcopy document provided by the aircraft manufacturer and carried during flight which lists checklists and procedures for various emergency scenarios. Jarvis (2017) noted that the QRH created a "substantial and unnecessary risk to flight safety" (p. 5).

Jarvis (2017) identified several issues with the QRH. Its physical bulk (250 pages) makes it difficult to manipulate, often requiring the dedication of one hand while the pilot simultaneously controls the aircraft with the other - a task which frequently requires two hands. The layout of the QRH is confusing and creates ambiguity, as checklist items are written as a linear set of steps, even when there are conditional elements and decision points. Furthermore, long sentences and superfluous words create the risk of pilots losing their place in the checklist or omitting items, while also increasing the time taken to complete a checklist. Additionally, the font size is below the recommended minimum size, there is no use of colour to highlight important items, and there are large sections written in capital letters, which is more difficult to read than sentence case[5]. Flying the T-6 while referring to the QRH therefore requires the pilot to use only short glances, as longer glances at the checklist, combined with the unfavourable trim characteristics, have the potential to cause undesirable flight path deviations since aircraft attitude cannot be monitored at the same time (Jarvis, 2017).

Jarvis (2017) concluded that simply flying the T-6C requires heightened attention switching over that required in similar aircraft types, and that the QRH, when used, imposed additional unnecessary demands on the pilot's mental resources due to its design. In times of an emergency, when cognitive processes are already under strain, noting that this is exactly the time the QRH will be used, this additional unnecessary workload could compound and result in an infringement of safety margins. The following excerpts from Jarvis' (2017) report summarise the research problem.

> In a single pilot, manually flown aircraft, the checklist design should make every effort to minimize the checklist's competition for the same pilot attention and visual resource as the primary flying task. However little or no concession is made; the checklist demands a great deal of pilot attention, where it need not do if it were designed and

[5] See any of the references on checklist design for an argument against these aspects.

written more effectively (p. 5). It creates an overall latent failure that will appear to have been obvious after an accident, and is in fact already obvious in foresight (p.26).

2.8 Electronic Checklist Application Development

Jarvis (2017) recommended the development of an electronic checklist to address the identified issues of the QRH. T-6C pilots operated with an iPad Mini kneeboard to access aeronautical map and chart information. These iPads offered a natural opportunity to migrate the QRH into a more user-friendly electronic form. A project was established to develop an Electronic Emergency Checklist (EEC) which would mitigate as far as possible the risks identified for the QRH. A review of industry guidance for electronic checklists and Burian's (2006) design principles informed the design of the EEC. Due to the commercial agreement with the aircraft manufacturer, checklist content was unable to be altered. The overall aims for the EEC project were to remove the risk around the physical manipulation of the paper QRH, improve content design and presentation to pilots of critical in-flight information, and reduce in-flight workload to enable a safer and more efficient cockpit for T-6C pilots.

2.9 Research Aim

The aim of the current research is to investigate whether workload is lower and flying performance is improved, when using a digital checklist application created specifically for the T-6C aircraft (the EEC), compared with the paper-based QRH.

2.10 Hypotheses

To support the aim, four hypotheses were derived from the academic literature and Jarvis' (2017) human factors report:

- H_1: The time taken to locate a specific checklist will be faster using the EEC than with the QRH.
- H_2: The time taken to complete a checklist will be faster using the EEC than with the QRH.
- H_3: Flight path accuracy within specified tolerances will be better with the EEC than the QRH[6].
- H_4: Pilot workload will be lower using EEC than the QRH.

Additionally, a qualitative investigation into error occurrences during checklist execution was undertaken for the QRH and EEC.

In the introduction it was explained that the underlying theme of this research report is that as pilot workload is increased, flying performance decreases. A checklist that requires high workload to operate should result in a flying performance decrease, and a checklist which requires lower pilot

[6] It is assumed less attention switching will be required with EEC use, resulting in a more accurate flight path for EEC users.

workload to operate should result in a performance increase. H_1 to H_3 encompass flying performance, and workload relates to H_4 and the qualitative investigation.

3. Method

3.1 Design

This study used a mixed-methods design. The experimental part of this research used a quantitative between-subjects randomised design. The independent variable was checklist type (paper vs electronic), and dependent variables were time to find the checklist, time to complete the checklist, flight path accuracy, and workload. An investigation into checklist error occurrences formed the qualitative part of this study.

The RNZAF Directorate of Flight Testing provided guidance on the conduct of the experiment and on workload measures to be used. A large amount of guidance was also taken from Casner & Gore (2010), who provide a framework by which to conduct an experiment assessing workload through use of a flight simulator.

3.2 Participants

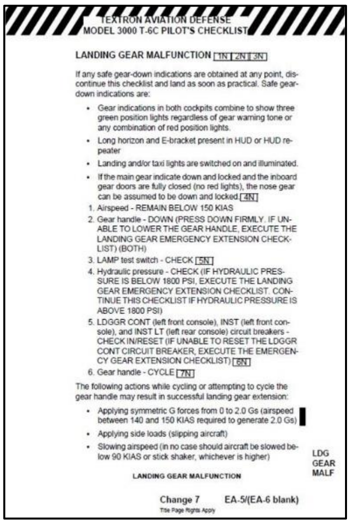
The participants were 20 pilots from across No. 14 Squadron, RNZAF, with a range of flying experience. Participant flying experience was reflected by pilot 'category' which is dictated by the number of flying hours. 'A-Category' is the most experienced, and 'student' is the least experienced. Seven of the pilots were 'A-Category' flight instructors, six were 'B-Category' flight instructors, three were 'C-Category' flight instructors, and four were student pilots who were undergoing pilot training. For the students, a requirement for participation was to have completed their instrument flying test. There were three females and 17 males with a mean age of 39.45 years (SD = 11.41, age range = 21 – 58). Flying hours ranged from 115 hours to 9500 hours (median = 3175, mean = 3656.5, SD = 2734).

3.3 Materials

3.3.1 Quick Reference Handbook (QRH)

The T-6C Quick Reference Handbook (QRH) (Figures 2 and 3) is a hardcopy document carried during flight which is used for in flight access to emergency checklists and other important operating data. Issues with the QRH are identified in section 2.7.

Figure 2: Pilot's Quick Reference Handbook (QRH)



TEXTRON AVIATION DEFENSE
MODEL 3000 T-6C PILOT'S CHECKLIST

LANDING GEAR MALFUNCTION [1N] [2N] [3N]

If any safe gear-down indications are obtained at any point, discontinue this checklist and land as soon as practical. Safe gear-down indications are:

- Gear indications in both cockpits combine to show three green position lights regardless of gear warning tone or any combination of red position lights.
- Long horizon and E-bracket present in HUD or HUD repeater
- Landing and/or taxi lights are switched on and illuminated.
- If the main gear indicate down and locked and the inboard gear doors are fully closed (no red lights), the nose gear can be assumed to be down and locked. [4N]

1. Airspeed - REMAIN BELOW 150 KIAS
2. Gear handle - DOWN (PRESS DOWN FIRMLY. IF UNABLE TO LOWER THE GEAR HANDLE, EXECUTE THE LANDING GEAR EMERGENCY EXTENSION CHECKLIST) (BOTH)
3. LAMP test switch - CHECK [5N]
4. Hydraulic pressure - CHECK (IF HYDRAULIC PRESSURE IS BELOW 1800 PSI, EXECUTE THE LANDING GEAR EMERGENCY EXTENSION CHECKLIST. CONTINUE THIS CHECKLIST IF HYDRAULIC PRESSURE IS ABOVE 1800 PSI)
5. LDGGR CONT (left front console), INST (left front console), and INST LT (left rear console) circuit breakers - CHECK IN/RESET (IF UNABLE TO RESET THE LDGGR CONT CIRCUIT BREAKER, EXECUTE THE EMERGENCY GEAR EXTENSION CHECKLIST) [6N]
6. Gear handle - CYCLE [7N]

The following actions while cycling or attempting to cycle the gear handle may result in successful landing gear extension:

- Applying symmetric G forces from 0 to 2.0 Gs (airspeed between 140 and 150 KIAS required to generate 2.0 Gs)
- Applying side loads (slipping aircraft)
- Slowing airspeed (in no case should aircraft be slowed below 90 KIAS or stick shaker, whichever is higher)

LDG GEAR MALF

LANDING GEAR MALFUNCTION

Change 7 EA-5/(EA-6 blank)
Title Page Rights Apply

Figure 3: A checklist as it is displayed in the QRH

3.3.2 Electronic Emergency Checklist (EEC)

The industry review enabled the development of a concept design which was refined and provided to a commercial app developer to create. The development of the EEC application took approximately four months. The fundamental design principle revolved around displaying flight critical information cleanly and simply, and the EEC sought to capitalise on the advantages an electronic platform could provide over a paper-based QRH. When initiating a checklist with the EEC, for example, the user is presented with a set of steps written in white text, with the first item enclosed within a magenta box, denoting the 'active' checklist item. As each checklist step is completed, the operator double taps the checklist item, which turns the step from white to green and causes the magenta box to highlight the next item in the sequence. The magenta box enables the operator to quickly find their place again after looking away. An undo function is available by double tapping a completed step, but steps cannot be completed or undone out of sequence.

Additionally, expanded information under 'notes' in the checklist can be collapsed once read to reduce checklist bulk by tapping the white 'note' box. Another feature of the EEC is the ability to direct a user down the appropriate decision tree when presented with a choice. The operator selects an option, and the checklist highlights the next step in the decision pathway. The EEC also incorporates an interactive menu, enabling quick and easy navigation to desired checklists, either through a contents page or a Crew Alerting System caption index, and a coordinate system for cockpit circuit breakers, enabling faster location. Once all items have been actioned, the words ’checklist complete‘ are displayed in green font. Several of these features are shown in Figures 4 and 5.

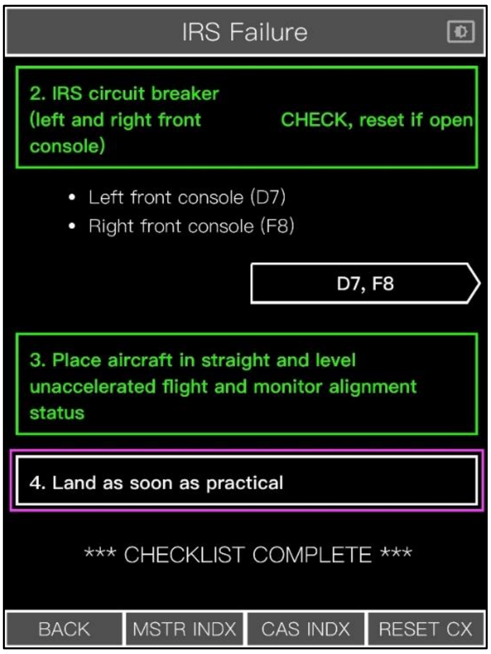


Figure 4: Electronic Emergency Checklist

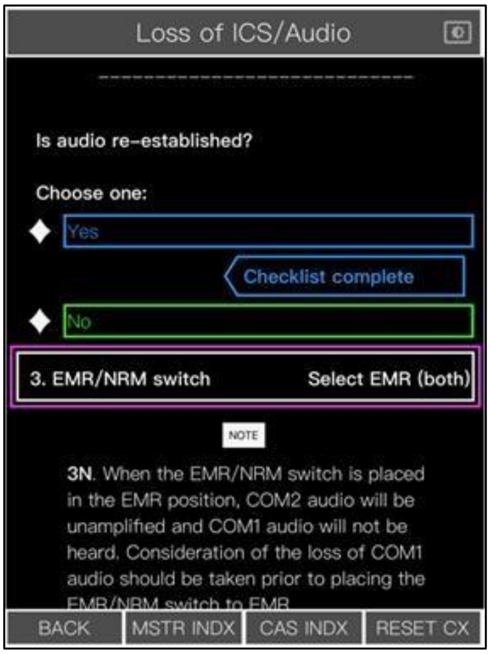


Figure 5: Electronic Emergency Checklist

3.3.3 Bedford Rating Scale

The Bedford Rating Scale (BRS) (Roscoe, 1984), developed for pilots, is a unidimensional workload rating scale designed to identify an operator’s spare mental capacity when completing a task. Workload is assessed using a hierarchical decision tree that guides the participant through a ten-point rating scale, with ten reflecting the highest workload and one the lowest. Each point is accompanied by a descriptor of the associated level of workload. It has been widely used to evaluate pilot workload in military and civilian settings, in both simulated and actual flight (Hart & Wickens, 1990). The BRS is shown at Appendix A.

3.3.4 Modified Cooper Harper Scale

The Modified Cooper Harper Scale (MCHS) was adapted by Wierwille & Casali (1983) from the Cooper Harper scale (Cooper & Harper, 1969) to extend the workload assessment capabilities of the Cooper Harper Scale to situations and tasks outside the psychomotor-aircraft piloting domain. Workload is again assessed using a ten-point rating scale, with ten reflecting the highest workload and one the lowest. The MCHS is a valid and statistically reliable indicator of overall mental workload (Gawron, 2000), and has been found to be sensitive to workload variations in simulated flight (Hart & Wickens, 1990). The MCHS is shown at Appendix B.

3.3.5 Flight Simulator

The T-6C flight simulator (Figures 6 and 7) is a fixed-base, scale mock-up of the front cockpit and replicates all controls and switches. The cockpit mock up is centred inside a visual dome on which an image is projected, providing for a 270 degree lateral, 43 degrees up and 28 degrees down image. A motion seat provides kinaesthetic and somatic feedback to the operator using algorithms designed to replicate aircraft movement. The simulator is controlled from an Instructor Operating Station (IOS) (Figure 8), which enables the simulator operator to input emergencies, reposition the aircraft, and control many other standard flight simulator functions. The IOS functions are also replicated on a portable tablet which can be used next to, or within the cockpit.

Figure 6: T-6C Flight Simulator

Figure 7: T-6C Flight Simulator

Figure 8: Instructor Operator Station

3.3.6 Simulator Playback Station

Data recorded from the simulator can be displayed at the playback station. The simulator automatically records a large number of parameters, which can be displayed individually or overlaid. To obtain a value at each data point, the station operator is required to hover the mouse pointer over the relevant point and read the value off the screen. Figure 9 shows the output format.

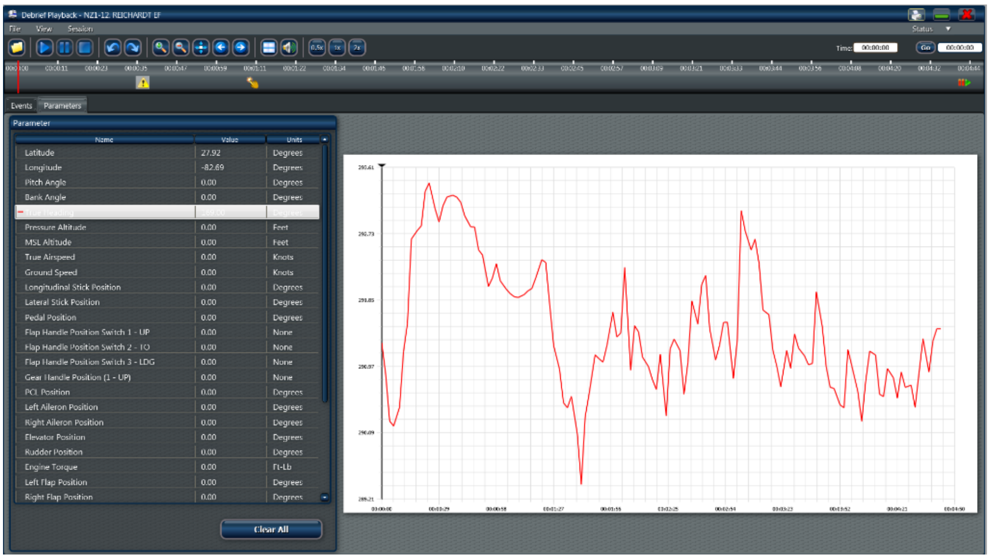


Figure 9: Simulator Output. This is a display of a heading plot. Time is on the x-axis and heading (in degrees) is on the y-axis. The units on the y-axis change depending on the parameter displayed. Altitude, for example would be displayed with 'feet' on the y-axis.

3.4 Dependent Variables

For this experiment, the primary task was operating the checklist and the secondary task was flying the aircraft. The secondary task of flying the aircraft is necessary, as without this task the operator is simply reading and operating the checklist with no other distractions – a situation in which useful data is unlikely to arise.

3.4.1 Checklist Find and Completion Time

Measuring the speed and accuracy at which an operator can perform a task is one of the simplest measures of performance. As workload increases, any performance decrease is likely to result in longer task completion times (Casner & Gore, 2010). For this experiment, it was predicted that the difficulty in operating the QRH would manifest itself firstly in the time taken to locate a particular checklist during an emergency, and secondly that it would take longer to complete the checklist once found. Participants were therefore timed on two measures. The first measure was the time taken to find the correct checklist (called 'checklist find time'), and the second measure was the time taken to complete the checklist (called 'checklist completion time').

A stopwatch was used to record checklist find time, as suggested by Casner & Gore (2010). The stopwatch was started when the participant physically lifted the QRH from the cockpit storage or selected the application icon on the EEC. The stopwatch was stopped when the participant found the correct checklist. Timing was conducted by two assessors, and the times were averaged post session.

Checklist completion time was calculated with the 'mark' function of the IOS. The 'mark' button was pressed when the stopwatch was started and at the completion of the scenario it was pressed again. A predetermined condition was the trigger for completion of the scenario (see section 3.5.2.1 and 3.5.2.2). The time difference between the 'marks', minus the checklist find time, was the checklist completion time.

3.4.2 Flight Path Accuracy

Jarvis (2017) observed that overly long glances at the QRH cause undesirable flight path deviations as aircraft attitude cannot be monitored at the same time, and that there is a heightened potential for 'upset attitudes' due to the continual attention switching required when using the QRH. Johannsen & Rouse (1983) used attitude deviations (such as pitch and roll angles), and frequency of tolerance exceedances as measures of performance. This principle has been applied to this experiment also.

Since the T-6C is particularly sensitive in pitch, this was the most obvious measure for attitude deviations. However, since the aircraft has no autopilot, if not trimmed correctly in roll and yaw, deviations would also manifest themselves in the roll and yaw axes during times of diverted attention. Altitude and heading deviations naturally result from movement in the three axes and so data from all three axes was collected to assess aircraft flight path.

The flight path accuracy dependent variable is made up of flight path exceedances (consisting of attitude, heading and altitude deviations outside specified tolerances), and flight path range (the maximum and minimum values for each flight path parameter). The numerous parameters within this dependent variable reflect the particular focus that was given to flight path accuracy, since it was anticipated that this dependent variable would best manifest the differences in pilot performance between checklist types.

3.4.2.1 Flight Path Exceedances

Heading, altitude, pitch, and roll parameters from the two scenarios were obtained from the respective plots at the playback station post-session. The design of scenario one enabled only altitude data to be recorded. Scenario two enabled all flight path parameters to be observed. Flight path tolerances were as follows:

- Pitch attitudes in excess of +/- 1 degree
- Pitch attitudes in excess of +/- 0.5 degrees[7]
- Roll attitudes in excess of +/- 5 degrees
- Altitude deviations in excess of +/- 100 feet
- Heading deviations in excess of +/- 5 degrees

The limits for altitude and heading were chosen as they demonstrate the flight test accuracy to which a pilot is expected to fly while operating with sole reference to instruments. This approach was also taken by Haddock & Beckman (2015) in a study of pilot workload during the approach phase of flight.

3.4.2.2 Flight Path Range

The highest and lowest values and the range of altitude, heading, pitch and roll were recorded. The data from these plots were obtained post-session from the recording and playback station.

3.4.3 Workload

The Bedford Rating Scale (Roscoe, 1984) and the Modified Cooper Harper Scale (Wierwille & Casali, 1983) were used. Both scales were administered to participants at the completion of each scenario, in line with Casner & Gore's (2010) suggested approach for a workload experiment.

3.4.4 Error Occurrences in Checklist Execution

Jarvis (2017) observed specific errors that occurred when using the QRH. These were loss of place, item omission, and incorrect actions. These three items were also categorised by Boorman (2001) as error modes associated with paper checklists. Along with a miscellaneous error category, these would initially form the four error categories to be investigated. To alert the assessors of errors, and to enable the assessors to follow through the scenario without interrupting, participants were asked to read the steps out aloud as they worked through the checklist. Casner & Gore (2010) suggested that errors are recorded by the researcher during the scenario without interrupting the participant, and this suggestion was adhered to. Interviews with the participants post-session, combined with assessor notes, formed the basis for the qualitative data. Originally, a thematic analysis was planned to be conducted on the miscellaneous error category only.

This approach was trialled in a pilot test (see section 3.5.5) where it was found that it was difficult to record errors in the four pre-planned error categories with any accuracy, even with the participant

[7] During data analysis, it became apparent that the resolution for the one degree pitch data was not adequate as there were many smaller deviations less than one degree which weren't accounted for, but which seemed important to include. To account for the smaller deviations, a 0.5 degree pitch parameter was added during the analysis phase.

speaking aloud. The error categories were subjective. For example, one assessor could categorise an error as item omission, while the other assessor may categorise the same error as a loss of place. It was decided, therefore, that all errors would be recorded individually along with a brief description, to be analysed through thematic analysis at the completion of testing.

3.5 Procedure

3.5.1 Participant Allocation

Participants were randomised within their experience categories and allocated to either the QRH or EEC group. Randomising within experience categories ensured roughly equal experience levels in both groups, and provided for a more representative sample than a simple randomisation of all participants where it would have been possible to randomly allocate all the experienced pilots in one group. Random allocation was achieved by drawing names out of a box. There were ten participants in each group.

Each group used their respective checklist types for both scenarios (i.e., the EEC group only used the EEC and the QRH group only used the QRH). A between subjects design was used to negate carry over effects and participant expectancy effect (although in principle this may not have been completely effective [see section 5.5]). Additionally, despite attempts to allocate squadron personnel to the study in advance, short notice absences (for other duties, sickness, etc.) meant that not all originally planned participants took part in the experiment and others were asked to fill in at short notice. Nevertheless, there remained equal experience levels spread between the two groups.

3.5.2 Simulator Sessions

Each simulator session consisted of two emergency scenarios and lasted for approximately 45 minutes. Participants were asked to attend the sessions in their flying gear, which consisted of ‘g’-pants and a harness, but no helmet. Helmets were omitted due to the generally poor simulator helmet communications system in the simulator, making it hard for the assessors to hear the participant. While it could be argued that the absence of helmets reduced the validity of the test, effective communication with participants was deemed more important. Participants wore a David Clark headset in place of a helmet.

Immediately prior to the simulator session, participants were given a brief covering the broad reasons for the experiment and how it would run, followed by a short training course on their respective checklist type. Despite the QRH participants having used the paper checklist previously, to avoid the possibility of lack of training arising as a confounding variable (where one group is briefed on their checklist type while the other isn’t), QRH users were still briefed on its use. At the completion of the brief, participants were asked to sign a consent form indicating that they were participating voluntarily and were also advised that they could withdraw from the study at any time.

Once the participant was seated in the simulator, they were given two minutes 'familiarisation time' with their checklist type. At the completion of the two minutes the checklist was reset (stowed - in the case of the QRH, or exited - in the case of the EEC), and the participant was given a brief for the first emergency scenario.

At the completion of each emergency scenario participants were asked to rate their respective checklist type on the Bedford and Modified Cooper Harper scales. This was followed by an unstructured interview, with the participant being asked for general comments and observations on the use of their checklist type during the scenario.

Two assessors were used to facilitate each session and to mitigate rater bias. One assessor was located at the IOS and was responsible for setting up the scenarios and managing the emergency scenario. The second assessor was located at the cockpit and conducted the briefs prior to each scenario. The assessors changed positions each session to mitigate fatigue and complacency. Both assessors made error recordings and notes for each scenario which were compared and summarised post-session.

3.5.2.1 Scenario 1

Scenario One required participants to conduct the 'Landing Gear Malfunction' checklist. This checklist was chosen since it was specifically mentioned by Jarvis (2017) as embodying all the identified problems with the QRH. It is long, complicated, and confusing. Jarvis (2017) found that eight out of nine pilots surveyed followed this checklist to the incorrect conclusion, despite having read it completely. Furthermore, this checklist provided the chance for many of the features of the EEC to be tested, such as choice pathways and in-checklist links. This scenario was conducted in good weather with a clear visual horizon and no cloud.

In this scenario, the landing gear malfunction was caused by the failure of the landing gear control solenoid (an electrical malfunction in the landing gear circuit preventing the gear handle from sending an electrical signal to the landing gear valve). This particular landing gear malfunction provided a single decision path to checklist completion and provided for the checklist in its entirety to be carried out. Other landing gear malfunctions do not run the checklist to completion and could elicit different decision pathways depending on pilot experience.

The participant was given a scenario brief (see Appendix C), and once comfortable in their seat they were instructed to take-off and conduct a circuit for a touch and go. The landing gear malfunction was programmed once airborne and with the gear raised. When the participant attempted to lower the landing gear in preparation for landing, nothing happened, and the participant began troubleshooting. This scenario was completed on achieving safe gear down indications with the emergency gear extension system. Safe gear down indications were the predetermined condition for the 'mark' button to be pressed to obtain checklist completion time. On completion of scenario one, the checklist was restowed, the simulator was reset, and participants were given the brief for scenario two (see Appendix C).

3.5.2.2 Scenario 2

Scenario Two involved the failure of one of the two on-board avionics computers. This checklist was selected since it was simpler than that of scenario one: it was short, it had a linear progression, and it required the pulling and resetting of a circuit breaker (which requires the pilot to physically look for the appropriate circuit breaker[8]). The scenario was conducted in cloud with no visual horizon and required the participant to maintain a single heading and altitude while managing the emergency. Due to the altitude and heading constraint, this scenario enabled the gathering of more flight path data than the first scenario.

On initiation of scenario two, participants were allowed 30 seconds of straight and level flight to familiarize themselves with the environment, after which the malfunction was introduced. This malfunction caused the master caution light to illuminate with the associated caption on the in-flight display. The scenario was completed (and the 'mark' button pressed) when the participant stated, "checklist complete".

3.5.3 Ethical Approval

Ethical approval to conduct the study was obtained from both the New Zealand Defence Force Organisational Research Branch and from Massey University. Ethical approvals from both organisations, along with the participant consent form, are included in Appendix D.

NZDF Organisational Research identified two issues. The first issue was that students could be influenced to participate involuntarily due to the researcher being a current T-6C flight instructor. The second issue was that it could be seen as unfair to students who did not participate in the experiment, in that the student participants got more 'practice' time in flying than those who did not participate in the experiment. The first issue was addressed by advising participants that participation was voluntary and obtaining written consent. The second issue, in consultation with the Squadron

[8] Circuit breakers are located along the side panels of the cockpit, which are slightly below and behind the pilot. The cockpit of the T-6C restricts movement and it is difficult to contort the body to read the circuit breaker labels on the panel. Flight path deviations can occur when locating a circuit breaker as the pilot is trying to move around in the seat, with attention diverted from controlling the aircraft.

commander, was deemed to be of minimal consequence given the other classroom learning tools available to the students, such as procedural trainers and desktop computer flight simulators. Furthermore, it was reinforced to all participants that performance in the experiment would not be assessed outside the study environment and would have no effect on course grades or career progression.

The ethics review at Massey University concluded that this project was of low risk and could therefore be registered on the database of low-risk studies.

3.5.4 Sample Size Determination

An '*a priori* power analysis' was run with the program G*Power to determine sample size (Appendix E). With $\alpha = .05$ and a large effect size ($d = .8$), a minimum of 26 participants in each condition was sufficient for experimental power ≥.80 in a two-tailed independent samples *t*-test for the between-subjects comparisons. (Note: Due to resource constraints and personnel availability on the squadron, 20 pilots underwent the experiment).

3.5.5 Pilot Test

A pilot test was completed prior to the experiment. This involved a single participant from outside the study group who conducted both scenarios, using the QRH for one scenario and the EEC for the other. As their results would not form part of the findings, written consent was not obtained, however, the reasons for the testing were explained, and the participant was a willing volunteer. The pilot test was run to identify any unforeseen issues with the experiment. The issues mentioned in section 3.4.4 in recording and categorising errors were highlighted during this test. Otherwise, the pilot test confirmed that the scenarios and checklist progression would run as anticipated and that the experiment could be managed with two assessors.

4. Results

All quantitative data was tested for normality using the Shapiro-Wilk test. Where the data did not meet normality, it was analysed using the independent samples *t*-test and the non-parametric Mann-Whitney U test. When both tests were used, the results were compared. In all cases, the Mann-Whitney U results were similar to the *t*-test results, and therefore only *t*-test results are reported here. Specific data that did not meet the normality criteria are detailed in the appropriate sections of this chapter. Alpha was set at 0.05 for all statistical tests, and all tests conducted were two-tailed[9].

4.1 Checklist Find and Completion Times

Checklist find time and checklist completion time for both scenarios by type of checklist are shown in Tables 3-5 and Figure 10. Only scenario one completion time (Table 4) showed a value outside normality. Table 3 shows the minimum and maximum find times, mean, and standard deviations for the time taken to find the correct checklist.

Table 3

Time taken to find the correct checklist.

	Checklist Type	*N*	Min/Max find time (seconds)	Mean find time (seconds)	SD Find time (seconds)
Scenario 1	EEC	10	5/41	17.20	11.33
	QRH	10	30/78	56.20	16.25
Scenario 2	EEC	10	4/50	14	14.80
	QRH	10	40/138	68.60	31.85

The time to find the checklist was significantly faster with the EEC than with the QRH for both scenario one, $t(18) = 6.224$, $p<.001$, $d = 2.78$ and for scenario two, $t(12.71) = 4.916$, $p<.001$, $d = 2.20$.

Table 4 shows the time taken to complete the checklist after it had been found. The table provides the maximum and minimum completion time, along with the mean completion time and the standard deviation. Two EEC participant's data was invalid for the first scenario due to incorrect checklist actions and supervisor injects. Overall, the time taken to complete each checklist is similar between both groups.

[9] Although all hypotheses were one-tailed, all tests of inference were conducted two-tailed to allow the reporting of any tests that were significant, but in the opposite direction to that which was predicted.

Table 4

Time taken to complete the checklist.

	Checklist Type	N	Min/Max Completion time (seconds)	Mean Completion time (seconds)	SD completion time (seconds)
Scenario 1	EEC	8	498/868	701.25	146.78
	QRH	10	410/1636	905.6	399.11
Scenario 2	EEC	10	84/343	210.1	80.98
	QRH	10	115/306	203.9	52.03

Checklist completion time showed no statistical significance between the two checklist types for scenario one, $t(16) = 1.369$, $p=.19$, $d = 0.68$ and scenario two, $t(18) = 0.204$, $p=.84$, $d = 0.09$.

Table 5 and Figure 10 provide descriptive statistics of the mean total time for checklist completion (checklist find time + checklist completion time). Table 5 displays the mean total time and Figure 10 shows the proportion of mean checklist find time and mean completion time to mean total time.

Table 5

Mean total time for checklist completion (mean find time + mean completion time).

	Checklist Type	Mean Total Time (seconds)
Scenario 1	EEC	718.45
	QRH	961.8
Scenario 2	EEC	224.1
	QRH	272.5

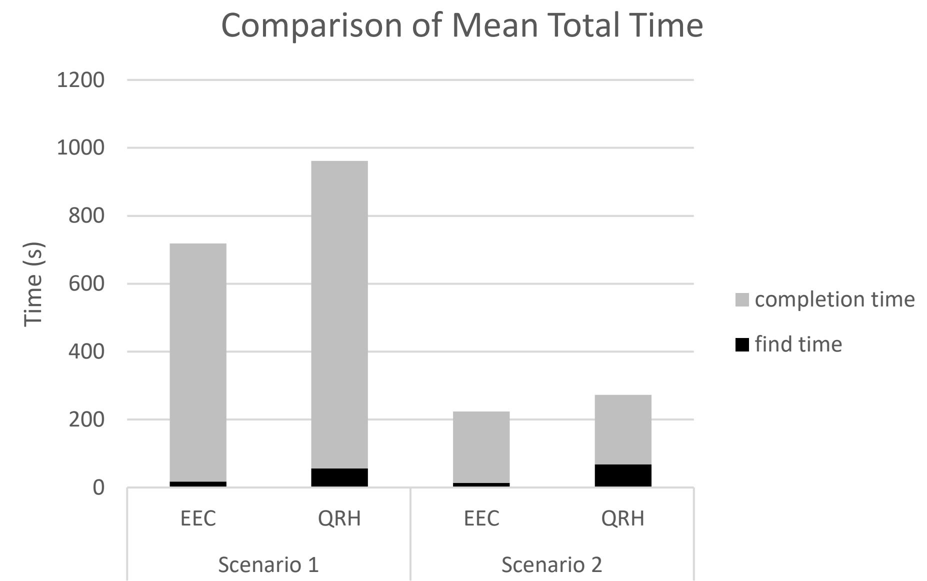


Figure 10: Comparison of mean find time and mean completion time to mean total time.

4.2 Flight Path Accuracy

4.2.1 Scenario One

The design of scenario one provided an opportunity to observe checklist use during a long and complicated procedure. Due to the nature of the scenario, however, in that the participant was only constrained in altitude and was free to otherwise manoeuvre during the conduct of the scenario, only altitude data was recorded. Scenario one results are shown at Tables 6 and 7 and in Figure 11. Descriptive statistics are displayed for each measure and are followed with their respective *t*-test results. All scenario one data met the criteria for normality under the Shapiro-Wilk test.

4.2.1.1 Scenario One Altitude Exceedances

Table 6 shows frequency of altitude exceedances (greater than +/- 100 feet) with each checklist type, including minimum and maximum frequency, mean and standard deviation for scenario one. Two EEC participants' data was invalid due to incorrect checklist actions and supervisor injects.

Table 6

Scenario one frequency of altitude exceedances (greater than +/- 100 feet).

Checklist Type	N	Min	Max	Mean	SD
EEC	8	1	6	3	1.77
QRH	10	0	7	2.9	2.42

The results suggest no significance in frequency of altitude deviations between checklist types, $t(16) = 0.097$, $p = .92$, $d = 0.05$).

4.2.1.2 Scenario One Altitude Range

Table 7 and Figure 11 show altitude range data for scenario one. Table 7 displays minimum and maximum altitudes and altitude range, along with mean altitude range and standard deviation. One QRH participant elected to remain at 2000 feet instead of 1200 feet for the scenario, and so their highest and lowest altitude data was an outlier and discounted (N=9). Figure 11 provides a graphical comparison of the scenario performance of the two checklist types.

Table 7

Scenario one altitude range. The datum altitude was 1200 feet.

Checklist Type	N	Min (feet)	Max (feet)	Range (feet)	Mean range (feet)	SD (feet)
EEC	8	914	1535	621	311.38	85.06
QRH	9	1033	1352	319	200.33	58.46

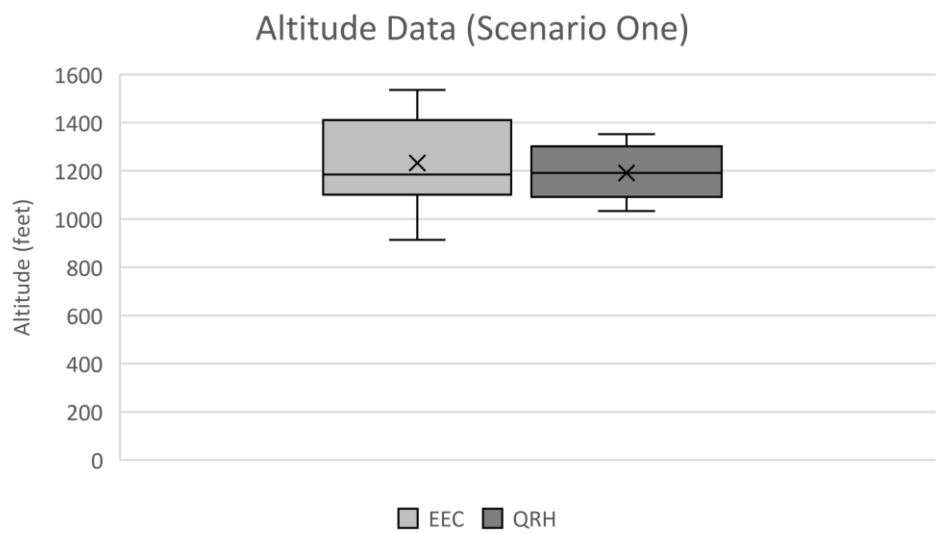


Figure 11: Scenario one altitude range data. The box and whisker plot displays minimum value (bottom line), first quartile, median value (middle line), mean (x), third quartile, and maximum value (top line).

Participants who used the QRH were able to maintain their altitude within a narrower band than those who used the EEC. The results show a statistically significant difference in favour of the QRH, $t(15) = 3.169$, $p = .006$, $d = 1.52$.

4.2.2 Scenario Two

Scenario two constrained participants to a specified altitude and heading, enabling a greater amount of flight path data to be gathered than in scenario one. All 20 participants provided valid data during this scenario. Tables 8-11, and Figures 12-15 display flight path data for scenario two.

Descriptive statistics are displayed for each measure and are followed with their respective *t*-test results. Altitude deviation, heading deviation and roll deviation frequency data showed values outside normality. As previously mentioned, only the *t*-test results are reported.

4.2.2.1 Scenario Two Flight Path Exceedances

Table 8 shows frequency of flight path exceedances and the associated limits in altitude, heading, pitch and roll parameters for each checklist type. Minimum and maximum frequencies are displayed, along with the mean values and standard deviation.

Table 8

Scenario two frequency of flight path exceedances.

	Checklist Type	Min	Max	Mean	SD
Altitude>+/-100ft	EEC	0	2	0.3	0.67
	QRH	0	2	0.2	0.63
Heading>+/- 5 deg	EEC	0	2	0.5	0.70
	QRH	0	2	0.6	0.70
Pitch >+/-0.5 deg	EEC	9	29	18.1	5.47
	QRH	7	24	18.1	5.02
Pitch >+/-1.0 deg	EEC	3	11	8.3	2.58
	QRH	1	13	7.1	4.36
Roll>+/-5 deg	EEC	0	5	2.6	1.43
	QRH	1	11	6	3.68

An independent samples *t*-test was run for each flight path parameter in Table 8 to compare the performance of each checklist. This analysis is displayed in Table 9. Altitude, heading, and pitch show no statistical significance between checklist type, while the roll parameter shows a statistically significant result in favour of the EEC.

Table 9

Statistical analysis for scenario two frequency of flight path deviations.

	t	df	*p*	*d*
Altitude>+/-100ft	0.342	18	.74	0.15
Heading>+/-5 deg	0.318	18	.75	0.14
Pitch>+/-0.5 deg	0	18	1	0
Pitch>+/-1.0 deg	0.749	14.63	.47	0.33
Roll>+/-5 deg	2.722	11.65	.02	1.22

4.2.2.2 Scenario Two Flight Path Range

Table 10 and Figures 12-15 show range data for scenario two flight path parameters. Table 10 displays minimum and maximum parameter values and range, along with mean range and standard deviation. Figures 12-15 provide a graphical comparison of the scenario performance of the two checklist types.

Table 10

Scenario two flight path range data. Datums for each parameter are as follows: Altitude = 4000 feet, Heading = 290 degrees[10], pitch = ~+0.5 degrees (corresponding to level flight), roll = 0 degrees (wings level). For pitch, a positive number describes a nose up pitch angle, a negative number is nose down. For roll, a positive number describes right bank, and a negative number is left bank.

	Checklist Type	Min	Max	Range	Mean Range	SD
Altitude (ft)	EEC	3883	4164	281	123.3	29.42
	QRH	3843	4171	328	118.6	47.19
Heading (deg)	EEC	285	298	13	6.4	1.43
	QRH	285	297	12	6.5	2.42
Pitch (deg)	EEC	-2.5	2.9	5.4	3.5	0.89
	QRH	-2.3	4.1	6.4	3.3	1.30
Roll (deg)	EEC	-11	12	23	13.67	3.61
	QRH	-12.7	13.8	26.5	15.82	4.79

[10] True heading. Magnetic headings are flown in practice, however the playback station only displayed true heading.

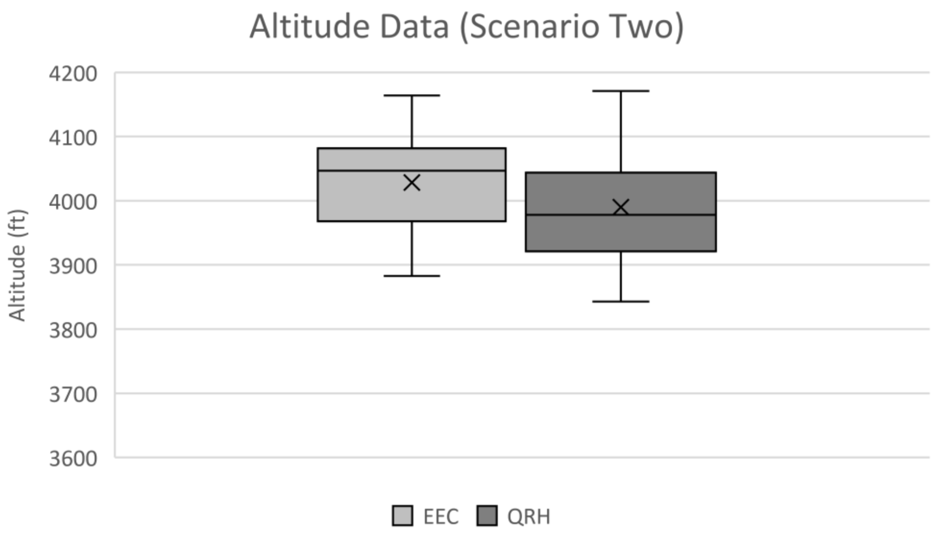


Figure 12: Scenario two altitude data. Datum altitude was 4000 ft.

Figure 13: Scenario two heading data. Datum heading was 290 degrees.

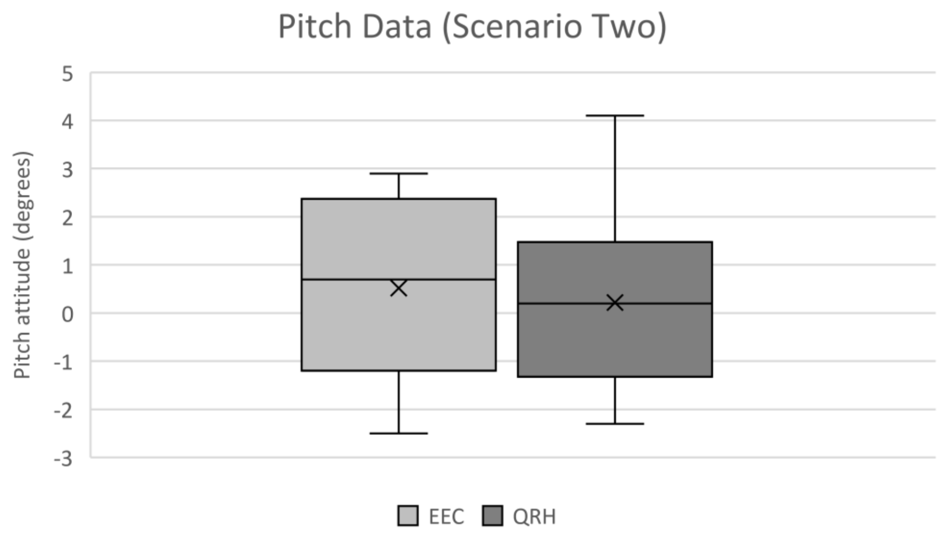


Figure 14: Scenario two pitch data. Datum pitch was ~+0.5 degrees (level flight).

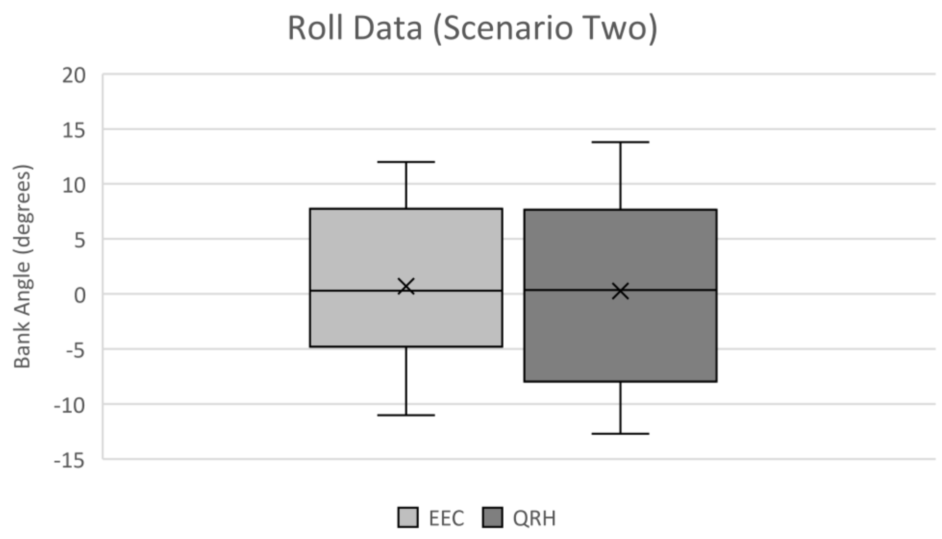


Figure 15: Scenario two roll data. Datum parameter was 0 degrees (wings level).

Table 11 shows the results of the independent samples *t*-test for scenario two flight path parameters. There is no statistical significance for any flight path range parameter between checklist types.

Table 11

Statistical analysis for scenario two range data.

	t	df	*p*	*d*
Altitude	0.267	18	.792	0.12
Heading	0.113	18	.912	0.05
Pitch	0.401	18	.693	0.18
Roll	1.134	18	.272	0.51

4.3 Workload

Bedford Rating Scale and Modified Cooper Harper Scale results are shown in Table 12 and Figure 16. Descriptive statistics are displayed for each measure. In keeping with the discussion in section 2.5.2, workload scores were treated as ordinal data and therefore statistical analysis was not conducted on workload results. However, basic data is provided in Table 12 and Figure 16 reflecting Annett's (2002) suggestion that simply showing that one result is preferred over the other is adequate for workload data. Overall, the EEC scored lower in workload than the QRH.

Table 12 displays the median, minimum and maximum values for workload measures over both scenarios. Figure 16 provides a graphical comparison of the scores for each checklist type and scenario.

Table 12

Bedford Rating Scale (BRS) and Modified Cooper Harper Scale (MCHS) scores.

		N	Median MCHS	Min/Max MCHS	Median BRS	Min/Max BRS
Scenario 1	EEC	8	2.5	1/4	3	2/3
	QRH	10	5	3/10	5	3/6
Scenario 2	EEC	10	2	1/4	3	1/3
	QRH	10	5	3/10	5	3/6

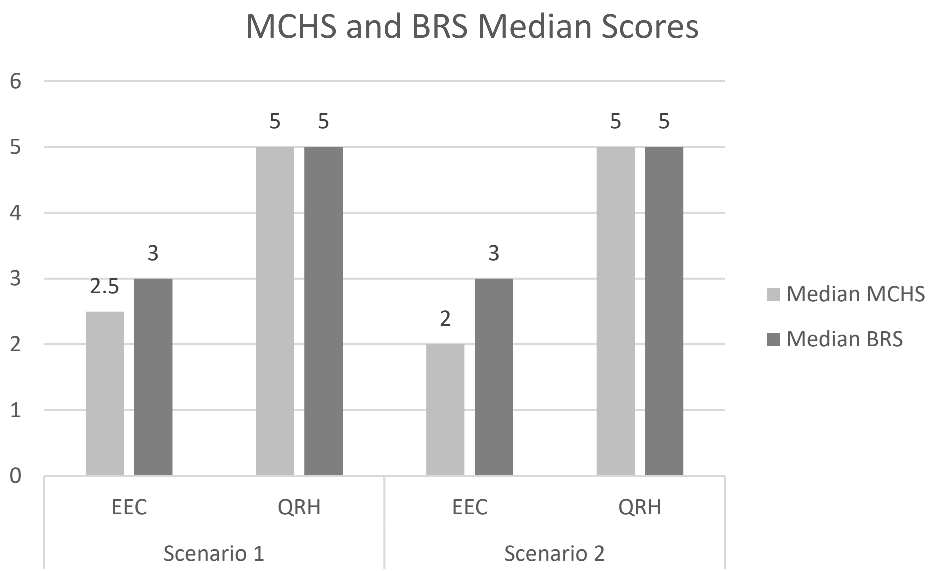


Figure 16: MCHS and BRS median scores.

4.4 Error Occurrences in Checklist Execution

The purpose of the qualitative aspect of this experiment was to determine if the frequency and type of error occurrences were affected by checklist type. Data for error occurrences was determined through qualitative analysis from interviews with participants post-session, and assessor observations during the simulator sessions. At the completion of the experiment, a thematic analysis was conducted following the six steps developed by Braun and Clarke (2006). While the initial objective of the investigation focused on error identification and categorisation, during the post-session interviews it became apparent that the topic was broader than focusing solely on errors. From the analysis, three main themes were generated roughly corresponding to Burian's (2006) emergency checklist design principles. These themes are checklist content, physical properties and usability, and checklist errors. Additionally, since the EEC was a safety-critical piece of equipment, a fourth theme was created specific to the EEC regarding the introduction of unintended human factor issues. The importance of avoiding introducing unintended human factor issues was also noted by Boorman (2001), during the introduction of the Boeing 777 ECL. The results of the thematic analysis are summarised in Table 13 and discussed in the following sections. A summary of the qualitative data can be found at Appendix F.

Table 13

Qualitative Analysis results for error occurrences in checklist execution using Braun and Clarke's (2006) six steps.

Subject	Themes	Codes
Error occurrences in checklist execution	Checklist Content	Poor/confusing wording Physical bulk/length of checklist
	Physical properties and usability	QRH is a loose item QRH is hard to navigate to find correct checklist
	Checklist Errors	Loss of place Item conducted out of sequence
	Unintended human factor Issues	Interaction issues Formatting issues Display issues

4.4.1 Checklist Content

Since the checklist content was unable to be altered due to commercial agreements, content issues were common to both checklist types as both the QRH and EEC were identically worded[11]. Checklist content was categorised into two subcategories: poor/confusing wording, and physical bulk/length of the individual checklist.

There were 20 examples of poor/confusing wording over the two OFT sessions. This equates to every person in the study identifying this issue. Examples consisted of poor conditional statements in the checklist and steps that were vague and easy to misread. Both these issues caused uncertainty in participants when deciding whether to carry out an action. In one case, a participant initiated the correct checklist, but as they progressed through the items, they directed themselves to the wrong checklist due to ambiguous checklist wording. In another case, a participant elected not to complete a checklist step due to their interpretation of the word 'reset' - even though the checklist explicitly defined the term. When taken in context with the particular checklist step, the word 'reset' became ambiguous, and so the participant omitted the step. Additionally, there were redundant and irrelevant terms in the checklist such as the use of 'inboard' gear door (discussed previously), and also a term specific to the US Navy which was not applicable to RNZAF operations. All these issues were

[11] This was noted by Carim et al. (2022) as a general disadvantage of regarding electronic checklists. Often, electronic checklists simply regurgitate the content of the paper-based QRHs, with any inherent limitations or failures in the procedures remaining present.

summarised succinctly by one participant who stated that "you have to really read it to try and understand what it is telling you".

There were six mentions of checklist 'bulk' over the course of the sessions. The landing gear scenario checklist in particular, described as "verbose" by one participant, is a long checklist covering several pages in the QRH. Both scenarios included a number of warnings, cautions and notes in the checklist, and there were comments on the sheer number of notes and other checklist content - so much so that participants would paraphrase, or skim read notes to complete the checklist more quickly so they could return to the primary task of flying the aircraft. In one case, a participant completely omitted reading a note due to its sheer volume, determining it was safer to concentrate on flying the aircraft rather than have their head down reading the large amount of words contained in the note. The checklist bulk and wording issues caused participants to raise concerns about the unnecessary and potentially dangerous length of time taken to complete a checklist. This reflects the point raised Carim et al., (2022) on the futility of a long checklist when there is not the time available to carry it out. More concerningly, the comments from participants suggest checklist neglect, which is worrisome, but particularly so in context of single pilot operations, when workload is already higher than in multicrew aircraft. As Degani & Weiner (1990) stated, "a long and detailed checklist is no guarantee of absolute safety. Indeed, it carries the risk that some pilots might choose not to use the checklist or conduct it poorly because of its length" (p. 20).

Checklist content issues are well documented in the literature. Indeed, Burian's (2006) three design principles are devoted almost entirely to remedying these issues. Too much detail in checklists can increase workload and become a precursor for pilot error (Burian, 2004; Federal Aviation administration, 1995). Pilots openly admitted that they would omit reading checklist content due to its bulk, and while this is concerning, it is not surprising, since humans often make decisions to avoid high workload. It should be kept in mind however, that this approach can have major negative effects on performance (Kool et al., 2010; Wickens, 2017). Furthermore, given that the primary role of the squadron is pilot training, the longer dwell times (or slower attention switching ability) of novice pilots when compared to experienced pilots (Tole et al., 1982) provides an additional safety concern regarding the checklist content. Due to the reluctance of the manufacturer to change the content unfortunately, the RNZAF must ensure that there is adequate training for pilots to understand the intent of the checklist.

4.4.2 Physical Properties and Usability

In contrast to the EEC, which is secured to the pilots leg with a Velcro strap, issues with QRH manipulation were numerous (28 QRH vs 2 EEC). Once removed from its stowage, the QRH is a loose item, and participants would either have it open and resting on their knee or would hold it in front of them with one hand. Both approaches had issues. For those who placed the checklist on their knee, often the checklist would begin to slide off their knee to the floor of the cockpit, causing the pilot to quickly stop what they were doing to catch QRH as it fell or moved, and resulting in a distraction from the primary flying task. Due to the way the pilot is strapped into the seat, once an object is on the cockpit floor it is impossible for the pilot to reach. Furthermore, once on the floor, the QRH becomes an unrestrained loose item which can, in the worst case, foul control cables and cause controllability problems. This issue has already been documented, and while in this case it did not directly affect the safety of the flight, the pilot did not realise the QRH had fallen to the floor until after landing. The alternative approach to placing it on the knee was to hold the QRH with one hand, which meant that participants were unable to easily manipulate switches in the cockpit, and some had to adjust their flying technique by switching hands on the control stick. Both cases (knee placement and holding with one hand) saw participants remove both hands from the control stick altogether to manipulate the QRH, leaving the aircraft not under direct control for short periods of time.

Despite a quick reference tab system, the QRH is loosely bound, and pages (made of waterproof paper) slide over one another when the QRH is held, often obscuring the quick reference tabs. Finding the correct checklist within the QRH often ended up with participants leafing through the document looking for the correct page number. To make matters worse, the QRH index is listed in alphabetical order from right to left - not left to right. Several participants mentioned that it was hard to find the correct checklist with the QRH. When combined with the content issues, another participant stated that it was "hard to get any 'flow' going with the QRH", while yet another said that if using the QRH, they would not refer to it at all for minor emergencies.

Additionally, the simulator environment was dim and to read the QRH effectively required artificial cockpit lighting directed onto the page. This caused issues for those pilots who chose to place the QRH on their knee, as it was further from the light source and harder to read.

Several participants commented on the fact that they were facing these difficulties in the simulator – a controlled, one 'g' environment with no turbulence or external distractions – and that if they had to use the QRH in the aircraft while in flight, the difficulties would be greatly magnified due to turbulence and other environmental factors.

In contrast to the QRH, the EEC was displayed on the iPad which was strapped to the participant's leg. The only manipulation issues with the EEC resulted from flying gloves not interacting well with the iPad touch screen (two occurrences), and complaints of a 'clunky' double tap to action checklist items[12].

4.4.3 Checklist Errors

Checklist errors were categorised as 'loss of place' and 'item conducted out of sequence'. The QRH scored highly in both 'loss of place' and 'item conducted out of sequence', while the EEC did not register any occurrences for either. This data is shown in Table 14.

Loss of place occurred from participants having to read the checklist warnings, cautions or notes[13] (W/C/N) that were located on a separate page, or from the frequent attention switching required to fly the aircraft. Referring to W/C/N required turning over pages to find the information, and then turning back to re-find the checklist and re-locate the appropriate checklist step. In one case, a participant inadvertently turned over two pages and struggled to re-find their place.

Task interruptions make it harder to resume the suspended task (Altman & Trafton, 2002). Therefore, 'item conducted out of sequence' was also likely due to pilot distraction from attention-switching. While similar to loss of place, this category involved a pilot conducting a checklist item without having completed the previous item or reading the incorrect W/C/N after turning the page. To mitigate this, many QRH participants used their thumb to keep their place in the checklist. When using the EEC, the application automatically tracked and indicated the pilots place through the magenta box. The higher workload values for the QRH were likely due to the prevalence of these error types.

Table 14

Frequency of errors and error category for each checklist type.

Error type	QRH	EEC
Loss of place	17	0
Item conducted out of sequence	8	0

4.4.4 EEC Human Factor Issues

A secondary goal of the experiment was to assess whether the EEC introduced any unintended human factors issues. Issues were identified from researcher observations and participant feedback, and

[12] If tapped too slowly, the checklist item would not mark as complete.
[13] See 'Definitions and Abbreviations'.

post-test were grouped into three categories: interaction issues, formatting issues, and display issues. Table 15 summarises the EEC human factor findings from the experiment.

Table 15

Human Factor issues identified from use of the EEC.

Human Factor issue	Description
Interaction	Forgetting to complete a checklist item. Single tapping instead of double tapping to complete item. Selecting magenta choice box instead of the choice diamonds.
Formatting	Incorrect action due to unfortunate line break. W/C/N after action is confusing.
Display	AFT circuit breaker panel was erroneously searched for. In the circuit breaker diagram the 'collared' circuit breaker depiction is confusing.

Interaction issues resulted from participants forgetting to double tap a checklist item to mark it complete, single tapping instead of double tapping to complete an item, and tapping the wrong area of the screen when presented with a decision point.

There were eight instances of a participant forgetting to double tap a checklist item. In these instances, the participant would progress a significant way through the checklist before realising that the previous items had not been marked as complete (turned green). Additionally, when a user is presented with a choice, if all items up to that point have not been marked complete, the choice function is unable to be actioned and the checklist progression stalls until all previous items have been double tapped complete. In other instances, participants tapped once to complete the checklist item before realising two taps were needed. Lastly, when presented with a decision point, the magenta box, which denotes the active item, encompasses the words "choose one:", prompting the user to select one of the choices adjacent to a white diamond (Figure 17). In several instances, the user would select the magenta box and get no response from the EEC, before realising the correct process to select the choice, which is to double tap the white diamond.

Formatting issues refer to the layout and display of EEC content. Formatting issues consisted of one occurrence of an undesirable line break and several occurrences relating to poor layout of warnings, cautions and notes. One participant provided a particularly good example of an unintended formatting consequence when they were directed to an incorrect checklist due to an unfortunate line break. The checklist item stated "…if unable to lower the gear handle, execute the landing gear emergency extension checklist." However, the EEC displayed "gear" and "handle" over two separate lines causing the participant to misread "if unable to lower the gear, execute the landing gear emergency extension checklist". In this scenario, the landing gear was stuck in the up position, but the gear handle was still able to be lowered. The participant identified that they could not lower the gear and then directed themselves to the incorrect checklist[14].

In the QRH, warnings, cautions and notes[15] are grouped on a separate page and referred to at the end of a checklist step if required[16]. Not every checklist step has an associated W/C/N. For example, in the QRH, "3N" at the end of a checklist step would prompt the user to refer to the relevant page and read note three before completing the checklist step. To provide consistency with the QRH, the EEC also listed W/C/N after the checklist item. However, this caused confusion, with some EEC users carrying out a checklist item and then reading the W/C/N information after the fact. During scenario one, for example, the EEC directs the user to test the landing gear indication lamps, with an associated note that explains the reason for conducting this check (i.e., to make sure a lamp is not blown). Inexperienced pilots would carry out the action, complete the item (double tap to turn it green), read the note, and then carry out the action again now knowing what to look for. In contrast, pilots had been trained to read any W/C/N in the QRH before they complete the step and so this issue did not occur with the QRH.

The last unintended human factor consequence regarded the display of the circuit breaker map (Figure 18). There were two issues identified. The first issue arose because the T-6C has a left and right circuit breaker panel in each cockpit (front and rear), with the circuit breaker map showing both cockpit panels on a single page. Often the checklist would give locations of front and rear circuit breakers and some EEC participants spent time erroneously looking for the rear cockpit circuit breakers (the rear cockpit circuit breaker panels cannot be reached from the front cockpit). This issue was not observed with QRH users, presumably because there is no picture in the QRH, rather the pilot is simply told which panel to look on (e.g., front left, rear right), and pilots know that the rear panels are inaccessible.

The second issue resulted from the EEC directing the user to a specific circuit breaker through a coordinate system. The EEC circuit breaker diagram, taken from the Aircraft Flight Manual, has bold

[14] This incident raises another checklist wording/content issue. Perhaps the word "lower" is not appropriate when referring to the landing gear handle and should be reserved for referring only to the landing gear itself.
[15] See 'Definitions and Abbreviations'.
[16] See Figure 3.

circles around the 'collared' circuit breakers[17]. In some instances, when referencing the circuit breaker map, participants were initially confused as they assumed the bold circles were automatically highlighted by the EEC to indicate the appropriate circuit breakers, despite the checklist coordinates referring to a different one.

Given the very short window (two minutes) that participants had to familiarise themselves with the EEC, and the fact that all participants could negotiate the application adequately (ignoring content issues), it is likely that a short training course would address all the human factors issues identified with the EEC. Nevertheless, if the opportunity arises to rectify deficiencies of the EEC through a software upgrade, then this should be considered.

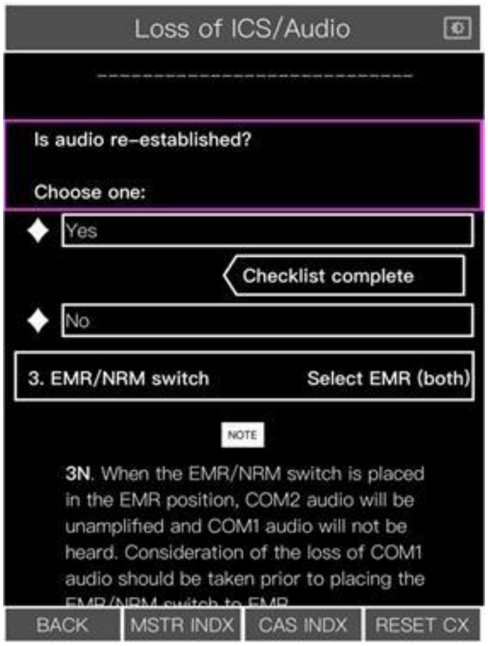


Figure 17: EEC choice function showing the magenta box and the choice options adjacent to the white diamonds. The user is unable to select the magenta box in this case and must select one of the options adjacent to the white diamonds.

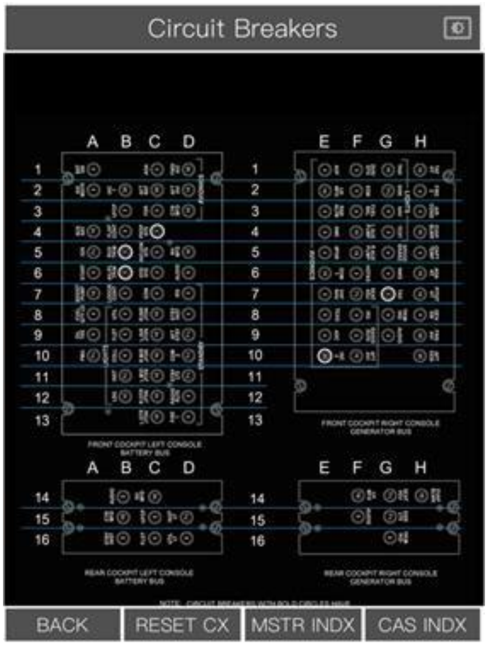


Figure 18: Circuit breaker diagram showing the EEC coordinate system. 'Collared' circuit breakers are indicated by a bold circle surrounding them. The front cockpit left and right panels are numbered 1-13. The rear cockpit panels are numbered 14-16. The rear cockpit panels cannot be reached from the front cockpit.

[17] Circuit breaker collars are used to raise flight critical circuit breakers above non-flight critical ones, enabling easy and quick identification by feel.

5. Discussion

5.1 Checklist Find and Completion Time

Checklist find time was faster using the EEC than with the QRH. This is almost certainly due to the EECs interactive menu, which likely reduces the attention switching demands and workload on the pilot by enabling access to any checklist with only a few button pushes. This is in contrast to the QRH, which requires significant physical manipulation to leaf through the document to locate the right page. The results provide evidence in support of H_1

The results for checklist completion time suggest no significance in completion time between the two checklist types, and therefore H_2 was not supported. One reason for the lack of significance in H_2 could be that checklist content is exactly the same between the two checklist types, and in the context of paper versus digital text, research suggests that there is no difference in reading speed and comprehension of text displayed on an iPad versus paper-based text (Sackstein et al., 2015; Stewart, 2012). A second reason is likely due to the location of the EEC. The EEC is poorly placed on the knee (Doherty, 2001), resulting in significant visual distance between the EEC and flight instruments, making attention switching difficult and inefficient, and extending completion time (Altman & Trafton, 2002). In contrast, many pilots in the QRH group were manipulating the QRH to be closer to the flight instruments to reduce the visual distances. A third reason could be because the experiment had less than the ideal number of participants. More participants may have resulted in statistical significance, considering that effect size was 0.68. A fourth reason, initially unforeseen by the researchers, may also help explain the similar completion time results. Scenario one involved a portion of the checklist where participants were required to conduct aircraft manoeuvring to try and force the landing gear down. Participants took a variety of approaches when conducting this manoeuvring portion of the 'landing gear malfunction' procedure. Some spent a long time manoeuvring, while others spent less, and some did not conduct the manoeuvring portion at all, having diagnosed the landing gear fault and correctly determining that manoeuvring would not assist in achieving a solution. Therefore, the time taken to complete scenario one varied widely. The author decided that the solution to this problem was to let the scenario play out for both groups of participants, on the assumption that with the roughly equal spread of experience between the two groups, this effect would cancel out. This issue did not occur in the second scenario as the scenario was much less dynamic, and participants were constrained in height and heading. Lastly, scenario one EEC results omitted two participants due to one occurrence of a participant completing the checklist out of sequence, and another occurrence due to a participant selecting the incorrect checklist, resulting in a supervisor inject part way through the scenario. This uneven grouping of participants between checklist types could have also influenced the results.

5.2 Flight Path Accuracy

Flight path accuracy was similar with both types of checklist, and there was no evidence of less attention switching for users of the EEC. The data therefore does not support H_3.

For scenario one, the results for frequency of altitude exceedances do not suggest a statistically significant difference between checklist types. In fact, for altitude range, the results show a statistically significant difference in favour of the QRH.

The manoeuvring portion of scenario one posed issues[18] for the researchers when monitoring altitude as the altitude exceedance parameters set by the researchers were purposely exceeded by the participant as they conducted their manoeuvring. Most participants requested, and were granted, a higher altitude to conduct the manoeuvring, following which they returned to the baseline altitude of 1200 feet. An attempt was made to mitigate this by bounding the manoeuvring portion with time "bookmarks" during the scenario. These bookmarks were displayed on the data plots at the playback station, and any altitude exceedances inside the bookmarks were ignored. The scenario was allowed to play out for both groups of participants on the assumption that, with the roughly equal spread of experience between the two groups, the effect would cancel out. This issue with altitude monitoring could account for the unexpected results in favour of the QRH.

In contrast, scenario two - by virtue of constraining participants to a height and a heading - enabled easier recording of flight path data and more flight path parameters to be observed. Frequency of altitude, heading and pitch exceedances show no statistical significance between checklist type, while frequency of roll deviations shows a statistically significant result in favour of the EEC. Flight path range data showed no significance across all parameters for scenario two.

During the data analysis, it was noted that in most cases standard deviations associated with EEC results were smaller than QRH standard deviations. Smaller standard deviations associated with the EEC would tend to indicate a more stable flight path overall, which could equate to more efficient attention switching in the EEC group. To investigate this further standard deviation scores for both exceedance and range data for both checklist types were standardised. Subsequent analysis again showed no statistical significance in the data. Overall, the results suggest that the use of the EEC does not improve flight path accuracy compared to the QRH.

---

[18] See Section 5.1.

These flight path results are in contrast to those found by Haddock & Beckman (2015), who compared the use of electronic charts displayed on an iPad with traditional paper charts on pilot performance during the approach phase of flight in single pilot aircraft. They found that pilots using electronic charts were able to remain within predefined flight path tolerances 70 to 80 percent of the time, compared to users of paper charts who remained in tolerances only 50 percent of the time. The study did not mention where the iPad was located in the cockpit, however.

Digital checklist location is an important consideration and should aim to minimise head down time and avoid interruptions, distractions, and spatial disorientation (Doherty, 2001). This principle was also identified by Jarvis (2017), who noted that the attention switching demanded by the QRH heightened the potential for 'upset attitudes'. Given the unergonomic placement of the EEC (on the pilots knee), it is likely that a similar amount of attention switching to the QRH was required to operate the EEC. This may also help explain the lack of significance in flight path data between checklist types, and the contrary results to those of Haddock & Beckman (2015).

This attention switching problem is particularly important in the case of novice pilots, according to (Tole et al., 1982), as their reserve of mental resources is less than that of experiences pilots. Novice pilots fixate longer on cockpit tasks which hinders their ability to switch attention effectively, making unintended deviations from flight path more likely as they take long glances at the checklist, regardless of checklist type. In the training environment especially, it is important that the EEC is ergonomically located.

5.3 Workload

Over both scenarios, participants rated the EEC much lower in workload than the QRH, suggesting that the EEC was easier to operate. The results therefore support H_4. These findings are consistent with Haddock & Beckman (2015), who found that participants who used an iPad for flight publication management and to display approach charts, reported significantly reduced stress levels and workload when compared to paper versions.

Several participants in the present experiment expected that workload would likely be higher in the aircraft[19]. From a physiological perspective, which was not investigated in this research report, this expectation is supported by Leino et al. (1995), who examined physiological measures among ten Finnish Air Force BA Hawk Mk 51 pilots and found that physiological measures of workload in real flight was significantly higher when compared to the same profile flown in a flight simulator. This finding is also supported by Wilson et al. (1987) in a study of A-7 attack aircraft pilots. However,

[19] See section 4.4.2

Magnusson (2002), in a study of five Swedish Air Force fighter pilots, disputes these finding, arguing that physiological reactions in the simulator are analogous to actual flight conditions. On the matter of subjective workload ratings, with which this research report is concerned, Denning (2003) found that subjective workload between simulated and real flight conditions was comparable.

While overall completion time for respective scenarios was similar amongst each group, suggesting similar workload for each checklist type, workload scores were very different for the two checklists. As has already been discussed, one reason which could be attributed to the similar completion time, especially in the case of scenario two, could be the poor placement of the EEC. A second reason could be attributed to the poor checklist content across both checklists, since many participants mentioned that the poor wording affected their workload. Both these reasons suggest that the lower workload score for the EEC was largely due to its improved interface over the QRH, and that the advantages provided by the improved interface are only secondary to better placement of either type of checklist, or a re-writing of the checklist content.

5.4 Error Occurrences in Checklist Execution

The qualitative data generated three themes: checklist content, physical properties and usability, and checklist errors. Additionally, since the EEC was a safety-critical piece of equipment, a fourth theme was created specific to the EEC regarding the introduction of unintended human factor issues.

Issues with checklist content was applicable to both checklist types. There were more issues related to physical manipulation with the QRH than with the EEC, since the EEC was strapped to the pilots knee. However, this placement of the EEC may have been detrimental to performance as it required pilots to switch attention over the large visual distance between the EEC and the flying instruments. This provides an opportunity for further investigation with a specific focus on checklist location. Several human factors issues were identified with the EEC which can be addressed through training or software upgrades.

5.5 Limitations

Several confounding variables could have affected the results of the experiment. First, the QRH was universally disliked amongst T-6 pilots and the airworthiness risk against it was widely known. Squadron pilots were aware of the intent of the experiment and almost certainly had biases towards a checklist solution that was not the QRH (participant expectancy effect). To attempt to mitigate this, information on the experiment was kept to a minimum, and all participants underwent the same induction process and respective checklist training briefs.

Second, the experiment did not obtain as many participants as the power analysis recommended. Only 20 participants were available out of the recommended 52. This was due to the small population of T-6C pilots. Additionally, several participants selected for the experiment were unavailable to attend and were replaced at short notice by available personnel. This meant that the study was conducted with a convenience sample.

Third, only two scenarios were used during the experiment. The squadron still had to maintain its regular training output while the experiment was running and therefore could only allocate a short window of time to the experiment. A greater number of scenarios would provide more data across all measures.

Fourth, 16 *t*-tests were conducted to analyse the results of the experiment, and consequently, the chance of Type 1 error is increased. This could explain why the roll parameter showed significance for scenario two.

Fifth, in the recording of altitude over both scenarios it was noted that often participants flew at an altitude slightly different to the desired baseline. For example, scenario one required participants to fly at 1200 feet, but many participants would accept some difference and would fly between 50-100 above the datum 1200 foot altitude. The same issue arose in scenario two with the 4000 foot datum altitude. To address this, the average height was estimated from the post-session data plots and any altitude deviations were assessed from this average height. For example, if a participant's average height was 4050 feet, the 100 foot altitude tolerance became 3950 and 4150 feet. Therefore, there is a margin of error on altitude data. In addition, the limitations of the recording and playback station meant that this average was assessed by eye and was not an exact science, providing further room for error. This problem did not occur with other flight path parameters.

Finally, and has already been alluded to, despite similar checklist completion times and flight path accuracy, one cannot rule out that the EEC in its current configuration (i.e., located on the pilot's knee) masked a result in favour of the EEC, as the knee placement still required the pilot to look down and away from the flying instruments while using it. Checklists located in obscure locations or on a kneeboard can pose a distraction to primary flying duties (Doherty, 2001), and this may be the case here. Given the sensitive pitch trim and distance between the visual references required to fly the aircraft and read the checklist, many pilots manipulated the QRH to be closer to the flight instruments, which could not be done with the EEC. Therefore, it is not possible to untangle fully the variance between the two groups in terms of the checklist interface versus the position of the checklist (i.e., was it the interface of the EEC or the position of the EEC that factored most heavily onto the results?).

5.6 Recommendations for future research

To fully determine the effect that checklist placement has on flying performance, future research would need to artificially standardise the placement of the QRH and EEC, or even compare two different positions (using the same checklist type) using a 2(EEC vs QRH) x 2(knee mounted vs control panel mounted) design.

Additionally, this experiment did not have access to eye tracking equipment and therefore the level of attention switching was inferred from the results. Future research should make use of eye tracking equipment to obtain quantitative data on attention switching. Future research may also benefit from the inclusion of more than two scenarios.

Finally, the aim of the qualitative aspect of this experiment was to formulate a hypothesis, which is that frequency and type of error occurrences are affected by checklist type. To test this hypothesis, further investigation should be undertaken with both the QRH and the EEC.

5.7 Implications of the Research

This research provides support for an iPad mount in front of the pilot in the T-6C. Additionally, this research provides further evidence that the T-6C checklist content is poorly written and can negatively impact flight safety and may assist in arguing for a content re-write.

From a wider perspective, most RNZAF pilots fly with a kneeboard and this research may be relevant for other aircraft types operated by the RNZAF, noting also that other aircraft types are flown by two pilots which may negate some of the findings in this research.

6. Conclusion

This research provided mixed results to support that flying performance is improved through the use of the EEC. Checklist find time was quicker with the EEC, but the EEC did not facilitate a faster checklist completion time or a more accurate flight path. There is evidence that workload is lower when using the EEC compared to the QRH. Subjective workload scores were overwhelmingly lower with the EEC compared with the QRH, suggesting the presentation and interface of the EEC was superior to the QRH. The qualitative investigation noted a reduction in errors and easier manipulation for the EEC, which also likely contributed to the lower subjective workload felt by participants. The collection of qualitative data enabled the generation of a hypothesis that frequency and type of error occurrences are affected by checklist type. The EEC introduced several unintended human factors issues. However, these can be addressed through training or software upgrades.

The EEC achieved all of the original aims of the checklist project: It removed the physical manipulation risk, improved content design and presentation of flight critical information, and it reduced workload. However, despite the improvements in the EEC interface over the QRH, this research suggests that checklist interface is a secondary consideration to greater factors at play: checklist location and content, and it is likely that the largest improvements in safety will be gained by better in-cockpit placement of the checklist or by a rewriting of the content.

Finally, it should be recalled from the opening pages of this paper that workload underscores all dependent variables observed in this experiment. A fundamental reason for the creation of the EEC was to reduce pilot workload, and despite the mixed results in flying performance, it can be concluded that the EEC is, in fact, an improvement over the QRH and provides for a safer cockpit environment in the T-6C.

References

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science*, 26(1), 39-83.

Annett, J. (2002*).* Subjective rating scales: science or art? *Ergonomics*, *45*(14), 966-987.

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, 22(4), 685-708.

Bailey Fausset, C., Weldon, E., Williams, L., Boyle, M., Cartwright, K., Robinson, M., Bosch, J., Crosland, M., & Meeks, B. (2020, December). Friendlier Skies: Development of Electronic Checklists for Emergency Procedures for the KC-135 Stratotanker. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 64, No. 1, pp. 139-143). CA: SAGE Publications.

Basil, M. D. (1994). Multiple resource theory I: Application to television viewing. *Communication Research*, *21*(2), 177-207.

Boorman, D. (2001). Safety benefits of electronic checklists: An analysis of commercial transport accidents. In *Proceedings of the 11th International Symposium on Aviation Psychology* (pp. 1-6). The Ohio State University.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

Bridger, R. (2017). *Introduction to Human Factors and Ergonomics*. CRC press.

Burian, B. K. (2004). Emergency and abnormal checklist design factors influencing flight crew response: A case study. In *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics 2004* (Vol. 1, p. 6).

Burian, B. K. (2006, October). Design guidance for emergency and abnormal checklists in aviation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 1, pp. 106-110). CA: Sage Publications.

Burian, B. K., Barshi, I., & Dismukes, K. (2005). *The challenge of aviation emergency and abnormal situations* (No. NASA/TM-2005-213462).

Burian, B. K., Dismukes, R. K., & Barshi, I. (2003, August). The emergency and abnormal situations project. In *Proceedings of the ISASI 2003 Conference*.

Cain, B. (2007). *A review of the mental workload literature*. Technical report, Defence Research and Development Toronto, Canada.

Carim, G., Torrisi-Steele, G., & Henriqson, E. (2022). Trapping Paper Checklists into Screens: How to Free the Resilience Capability of Digital Checklists for Emergency and Abnormal Situations. In *Resilience in a Digital Age* (pp. 271-284). Springer, Cham.

Casner, S. M., & Gore, B. F. (2010). *Measuring and evaluating workload: A primer*. NASA Technical Memorandum, 216395, 2010. CA: Ames Research Center.

Charlton, S. G. (2019). Measurement of cognitive states in test and evaluation. In *Handbook of Human Factors Testing and Evaluation* (pp. 97-126). CRC Press.

Civil Aviation Authority, United Kingdom (2000). *Guidance on the Design, Presentation and Use of Electronic Checklists*. TSO (The Stationary Office) on behalf of the UK Civil Aviation Authority

Cooper, G. E., & Harper, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities*. National Aeronautics and Space Administration.

Corwin, W. H., Sandry-Garza, D. L., Biferno, M. H., Boucek Jr, G. P., & Logan, A. L. (1989). *Assessment of crew workload measurement methods, techniques and procedures. Volume 1. Process, methods and results*.

Degani, A. (1992). *On the typography of flight-deck documentation* (No. A-93049).

Degani, A., & Earl, L. Wiener (1990): *Human Factors of Flight-Deck Checklists: The Normal Checklist*. National Aeronautics and Space Administration, Ames Research Center.

Degani, A., & Wiener, E. L. (1993). Cockpit checklists: Concepts, design, and use. *Human Factors*, *35*(2), 345-359.

Denning, T. V. (2003). *An experimental investigation of the differences in subjective pilot workload across simulated and real flight conditions*. Doctoral dissertation, Embry-Riddle Aeronautical University.

Diehl, A. E. (1991). Human performance and systems safety considerations in aviation mishaps. *The International Journal of Aviation Psychology*, *1*(2), 97-106.

Dismukes, K., & Berman, B. (2008). *Checklists, Monitoring, and Multitasking in Cockpit Operations*. FAA Research Review.

Dismukes, K., & Nowinski, J. (2007). Prospective memory, concurrent task management, and pilot error. In A. Kramer, D. Wiegmann, & A. Kirlik (Eds.), *Attention: From theory to practice* (pp.225–238). Oxford University Press.

Doherty, P. J. (2001). *Electronic Checklists on Multi-Purpose Displays: A Better Way for Fighter Pilots to Manage Information and Situational Awareness during Periods of High Workload*. Army Command and General Staff College. Fort Leavenworth, KS. School of Advanced Military Studies.

Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W. S., Wilson, J. R & van der Doelen, B. (2012). A strategy for human factors/ergonomics: developing the discipline and profession. *Ergonomics*, *55*(4), 377-395.

EASA. (2012). *Principles and guidelines relative to the design of checklists and working methods in the cockpit*. European Aviation Safety Agency, Safety Analysis and Research Department.

Epling, S. L. (2017). *Dual-tasking and multiple resources: the interference of cognitive and physical demands in real-world applications*. University of Canterbury (Dissertation).

Federal Aviation Administration. (1995). *Human performance considerations in the use and design of aircraft checklists*. US Department of Transportation.

Gawron, V. J. (2000). *Human performance measures handbook*. Lawrence Erlbaum Associates Publishers.

Haddock, K. N., & Beckman, W. S. (2015). The effect of electronic flight bag use on pilot performance during an instrument approach. *The Collegiate Aviation Review International*, 33(1).

Hales, B. M., & Pronovost, P. J. (2006). The checklist—a tool for error management and performance improvement. *Journal of Critical Care*, *21*(3), 231-235.

Hart, S.G., Wickens, C.D. (1990). Workload Assessment and Prediction. In: Booher, H.R. (eds) *Manprint* (pp. 257-296). Springer, Dordrecht. https://doi-org.ezproxy.massey.ac.nz/10.1007/978-94-009-0437-8_9

Hawkins, F. H., & Orlady, H. W. (2017). *Human Factors in Flight*. Routledge.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: when is one scale better than many? *Human Factors*, 35(4), 579-601.

Higgins, W. Y., & Boorman, D. J. (2016). An Analysis of the Effectiveness of Checklists when combined with Other Processes, Methods and Tools to Reduce Risk in High Hazard Activities. *Boeing Technical Journal*.

International Civil Aviation Organization (1998), *Human Factors Training Manual* (Doc 9683-AN/950). Montréal: ICAO

International Ergonomics Association (2022). *What is Ergonomics*? https://iea.cc/what-is-ergonomics/

Jarvis, S. (2017). *Technical Human Factors Review of the T-6 as a Pilot Training Platform.* Jarvis Bagshaw Ltd, UK.

Johannsen, G., & Rouse, W. B. (1983). *Studies of planning behavior of aircraft pilots in normal, abnormal, and emergency situations*. IEEE transactions on systems, man, and cybernetics, (3), 267-278.

Kantowitz, B. H. (2000). Attention and mental workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 44, No. 21, pp. 3-456). CA: SAGE Publications.

Kantowitz, B. H., & Casper, P. A. (2017). Human workload in aviation. In R. Dismukes (Ed.), *Human error in aviation* (pp. 123-153). Routledge. https://doi-org.ezproxy.massey.ac.nz/10.4324/9781315092898

Kavanagh, J. (2005). *Stress and Performance A Review of the Literature and its Applicability to the Military*. CA: RAND Corporation

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: general*, 139(4), 665.

Leino, T., Leppäluoto, J., Huttunen, P., Ruokonen, A., & Kuronen, P. (1995*).* Neuroendocrine responses to real and simulated BA Hawk MK 51 flight. *Aviation, space, and environmental medicine*, *66*(2), 108–113.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., & Linton, P. M. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies*. Technical report 851. United States Army Research Institute for the Behavioural and Social Sciences

Magnusson, S. (2002). Similarities and differences in psychophysiological reactions between simulated and real air-to-ground missions. *The International Journal of Aviation Psychology*, 12(1), 49-61.

Miller, S. (2001). *Workload measures*. National Advanced Driving Simulator. Iowa City, United States.

Myers III, P. L. (2016). Commercial aircraft electronic checklists: benefits and challenges (literature review). *International Journal of Aviation, Aeronautics, and Aerospace*, *3*(1), 1-10.

Orasanu, J. M., & Backer, P. (1996). Stress and Military. In J. Driskell & E. Salas (Eds.), *Stress and human performance* (pp. 89-125). Lawrence Erlbaum Associates.

Orlady, H. W., & Orlady, L. M. (1999). *Human Factors in Multi-Crew Flight Operations.* Aldershot: Ashgate Publishing Limited.

Roscoe, A. H. (1984). *Assessing pilot workload in flight*. Royal Aircraft Establishment Bedford (United Kingdom) Bedford United Kingdom.

Roscoe, A. H. (1987). *The Practical Assessment of Pilot Workload: Flight Mechanics Panel of AGARD*. Advisory Group for Aerospace Research and Development Neuilly-sur-Seine (France).

Ross, P. (2004). Human factors issues of the aircraft checklist. *Journal of Aviation/Aerospace Education & Research*, 13(2), 9-13.

Sackstein, S., Spark, L., & Jenkins, A. (2015). Are e-books effective tools for learning? Reading speed and comprehension: iPad vs. paper. *South African Journal of Education*, 35(4), 1-14.

Sirevaag, E. J., Kramer, A. F., Reisweber, C. D. W. M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, *36*(9), 1121-1140.

Snead, A. E., Militello, L. G., & Ritter, J. A. (2004, September). Electronic checklists in the context of B-2 pilot decision making. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 3, pp. 640-644). CA: SAGE Publications.

Stewart, S. M. (2012). *Reading in a technological world: Comparing the iPad to print* (Doctoral dissertation, Bowling Green State University).

Tole, J. R., Stephens, A. T., Harris, R. L., & Ephrath, A. R. (1982). Visual scanning behavior and mental workload in aircraft pilots. *Aviation, Space, and Environmental Medicine*, *53*(1), 54-61.

Turner, J. W., & Huntley Jr, M. S. (1991). *The use and design of flightcrew checklists and manuals*. John A Volpe National Transportation Systems Center.

Velázquez, J., Peck, A., & Sestak, T. (2015). Behavioral traps in crew-related aviation accidents. In *18th International Symposium on Aviation Psychology*. May 4-7, 2015, Dayton, Ohio, 372-377.

Wagner, D., Birt, J. A., Snyder, M. D., & Duncanson, J. P. (1996). *Human Factors Design Guide (HFDG) for Acquisition of Commercial off-the-shelf Subsystems, Non-Developmental Items, and Developmental Systems* (No. DOT/FAA/CT-96/1).

Warm, J. S., Matthews, G., & Finomore Jr, V. S. (2017). Vigilance, workload, and stress. In J. Szalma & P. Hancock (Eds.), *Performance Under Stress* (pp. 131-158). CRC Press. https://doi-org.ezproxy.massey.ac.nz/10.1201/9781315599946

Wickens, C. D. (1987, January). Attention in aviation. In *Proceedings of the Fourth International Symposium on Aviation Psychology* (pp. 602-608).

Wickens, C. D. (2017, June). Mental workload: assessment, prediction and consequences. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 18-29). Springer, Cham.

Wickens, C. D., & McCarley, J. S. (2019). *Applied attention theory*. CRC press.

Wiegmann, D. A., & Shappell, S. A. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system*. Ashgate.

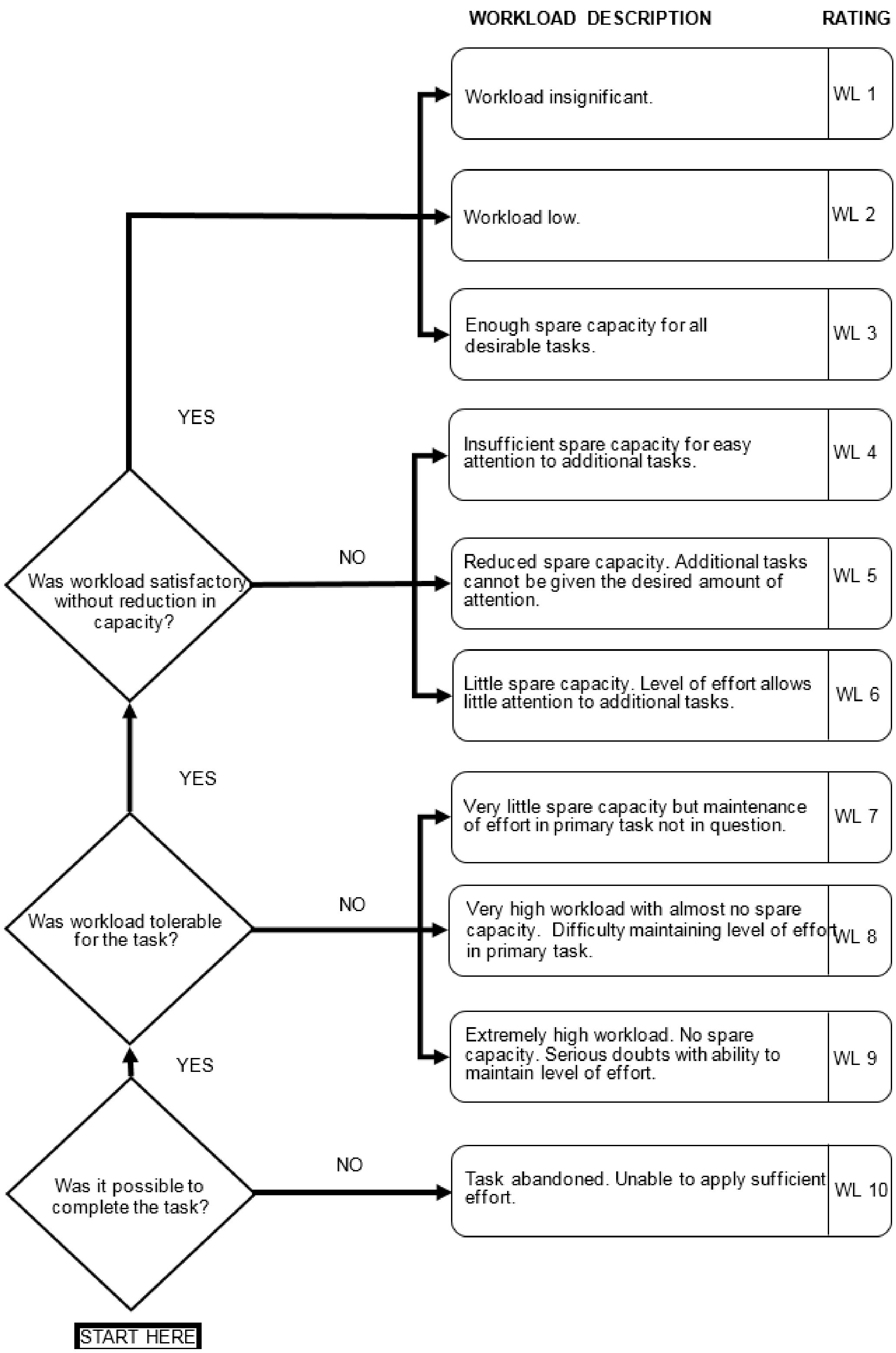
Wierwille, W. W., & Casali, J. G. (1983, October). A validated rating scale for global mental workload measurement applications. In *Proceedings of the human factors society annual meeting* (Vol. 27, No. 2, pp. 129-133). CA: Sage Publications.

Wilkins, S. A. (2018). Examining head-down time in transportation: case study in single-pilot general aviation operations. *Transportation research record*, *2672*(23), 137-145.

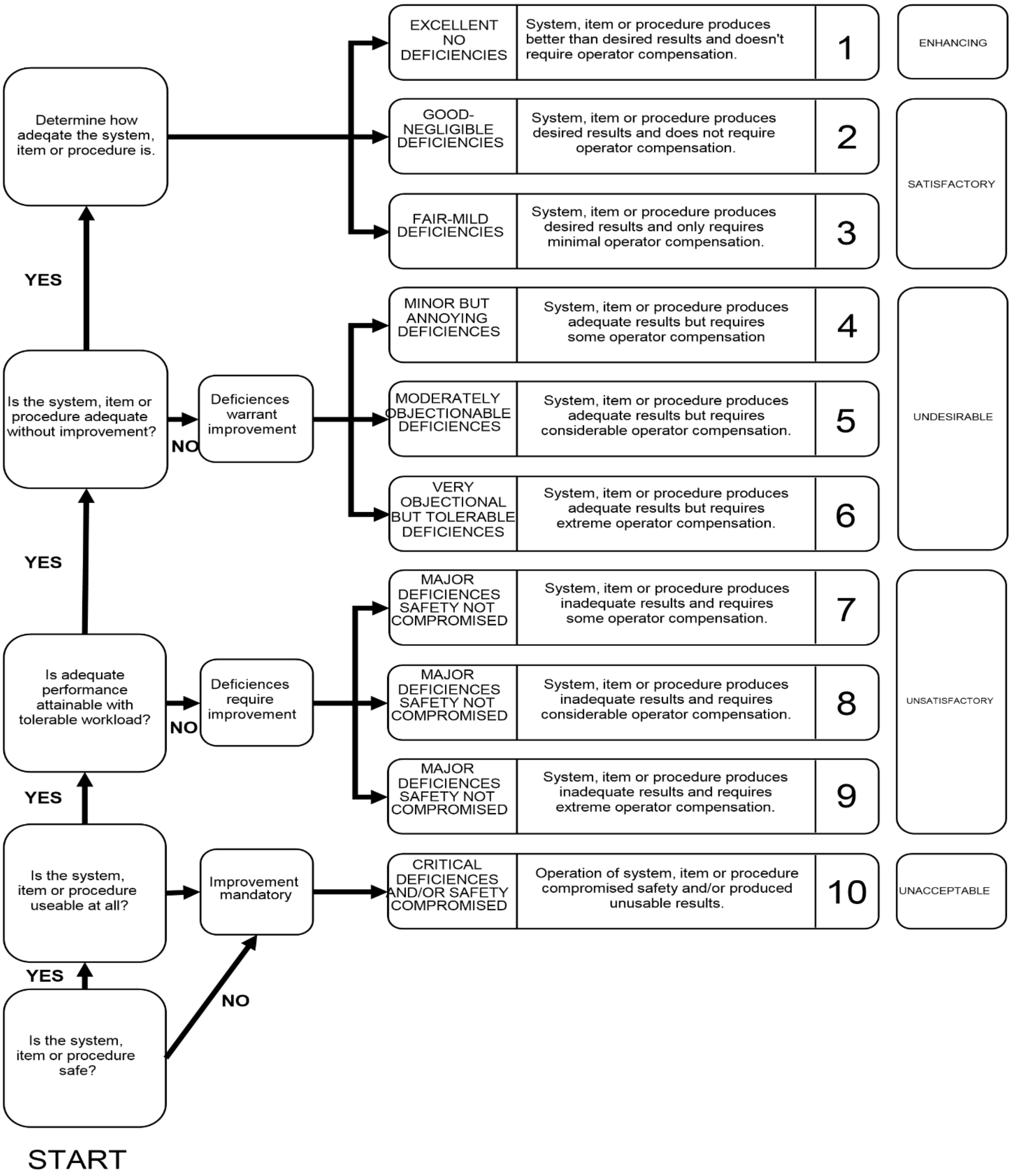
Wilson, G. F., Purvis, B., Skelly, J., Fullenkamp, P., & Davis, I. (1987, September). Physiological data used to measure pilot workload in actual flight and simulator conditions. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 31, No. 7, pp. 779-783). CA: Sage Publications

Wilson, J. R., & Rajan, J. A. (1995). Human-machine interfaces for systems control. In J. R. Wilson, and E. N. Corlett (Eds.), *Evaluation of human work: a practical ergonomics methodology* (pp. 357–405). London: Taylor & Francis.

Appendix A: Bedford Rating Scale



Appendix B: Modified Cooper Harper Scale



Appendix C: Scenario Briefs

Scenario 1: 3204 Landing Gear Control Valve Failure

Conditions:

Clear horizon, 80km visibility, nil cloud/wind, nil turbulence, QNH 1013, wind 270/10

Set up:

Threshold of runway 27, engine running, before take-off checks completed.

Brief:

"You are set up on the threshold of runway 27. The before take-off checklist has been completed and the aircraft is set up for take-off. Your clearance is to make a left turn after take-off and to conduct a normal circuit. Weather conditions are no wind or cloud and unlimited visibility. Once the scenario starts, manage the situation in accordance with clearance given to you by ATC (this is for validation purposes). 14 Squadron Ops is not available once the scenario starts. Lastly, this scenario is not assessed as part of any training courses and will not be discussed outside the study environment".

If you conduct an abnormal checklist, please:

- Read the checklist items aloud.
- Please state, "checklist complete" at the completion of any abnormal checklist.
- Notify supervisors of any errors or personal observations- provide explanation if needed at the end.

Any questions?

Once I state, "you have control", conduct a departure brief if you would like and then take-off and fly a left-hand circuit RWY 27.

Constraints:

- 14 Squadron Ops not available to assist.
- Tower clearance – Maintain 1200ft.

Procedure:

1. Crosswind turn - activate 3204 'landing gear control valve failure'.
2. Start recording at initiation of the checklist.
3. Stop recording on all gear down (3 greens) via emergency extension.

BRS and MCHS objectives

- Use the BRS to describe your workload while using your checklist type.
- Use the MCHS to describe your workload while using your checklist type.

Scenario 2: 3419 IAC 1 FAIL

Conditions:

Overcast ceiling 2000ft, cloud tops 10,000ft, wind 090/10

Set up:

4000ft early downwind vectors for ILS/DME runway 09, HDG 270, 180kt

Brief:

"You are early downwind for runway 09 at 4000ft on heading 270. Once off flight freeze, you'll have 30 seconds to stabilize the aircraft and familiarize yourself with the environment, and then the scenario will begin. ATC will clear you as required throughout the scenario. 14 Squadron Ops is not available once the scenario starts. Lastly, this scenario is not assessed as part of any training courses and will not be discussed outside the study environment".

If you conduct an emergency checklist, please:

- Read the checklist items aloud.
- Please state, "checklist complete" at the completion of any abnormal checklist.
- Notify supervisors of any errors or personal observations- provide explanation if needed at the end.

Any questions?

Once I state, "you have control", maintain heading 270.

Constraints:

- 14 Squadron Operations not available to assist.
- ATC clearance – Maintain 4000ft on heading 270.

Procedure:

1. Once 30 seconds has elapsed (and participant is stable) - activate 3419 'IAC 1 FAIL' malfunction.
2. Start recording at initiation of the checklist (reaching for paper checklist or opening EFB application).
3. Stop recording when "checklist complete" is stated by participant.

BRS and MCHS objectives

- Use the BRS to describe workload while using your checklist type.
- Use the MCHS to describe your workload while using your checklist type.

Appendix D: Ethical Approval

Kia ora,

Link to the application
HoU Review Group

Ethics Notification Number: 4000025393
Title: T-6C Electronic Checklist Application Project

Thank you for your notification which you have assessed as low risk.

Your project has been recorded in our database for inclusion in the Annual Report of the Massey University Human Ethics Committee.

The low risk notification for this project is valid for a maximum of three years.

Please notify me if situations subsequently occur which cause you to reconsider your initial ethical analysis that it is safe to proceed without approval by one of the University's Human Ethics Committees.

Please note that travel undertaken by students must be approved by the supervisor and the relevant Pro Vice-Chancellor and be in accordance with the Policy and Procedures for Course-Related Student Travel Overseas. In addition, the supervisor must advise the University's Insurance Officer.

A reminder to include the following statement on all public documents:

"This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named in this document are responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you want to raise with someone other than the researcher(s), please contact Professor Craig Johnson, Director (Research Ethics), email humanethics@massey.ac.nz."

Please note that if a sponsoring organisation, funding authority or a journal in which you wish to publish require evidence of committee approval (with an approval number), you will have to complete the application form again answering yes to the publication question to provide more information to go before one of the University's Human Ethics Committees. You should also note that such an approval can only be provided prior to the commencement of the research.

You are reminded that staff researchers and supervisors are fully responsible for ensuring that the information in the low risk notification has met the requirements and guidelines for submission of a low risk notification.

If you wish to print an official copy of this letter:

1. Please login to the RIMS system (https://rme.massey.ac.nz).
2. In the Ethics menu, select Ethics Applications.
3. Using the Advanced search with appropriate criteria to find only this application.
4. With the application on the Results tab, select Reports from the toolbar.
5. Select the "Human Ethics - Low Risk Notification Letter" link, this will open the report viewer.
6. Select the application code from the Report Parameters dropdown and submit. You can then select an export option from the top toolbar (Print, Save).

Yours sincerely
Professor Craig Johnson
Chair, Human Ethics Chairs' Committee and
Director (Research Ethics)

HEADQUARTERS NEW ZEALAND DEFENCE FORCE
Organisational Research

MINUTE

5000/PB/5/3

November 2021

ACOD

7.12.21

APPROVAL TO CONDUCT RESEARCH: EVALUATION OF T-6C ELECTRONIC CHECKLIST (ORG RESEARCH 2021/34)

References:
A. DFO 3, Chap 5, Part 14: Authority to Conduct Human-related Research

Background

1. In accordance with Ref A, SQNLDR Peter Barron, has requested approval to conduct research to evaluate the use of an electronic checklist for pilots.

2. This project aims to determine if the use of a T-6C Electronic Checklist application can address the human factor deficiencies of the T-6C Quick Reference Handbook (QRH).

3. This project is expected to benefit the NZDF by eliminating the airworthiness risk against the Texan checklist. The electronic checklist for the T-6C was initiated to address human factor issues with the paper copy QRH. The project developed a purpose built application for use inflight to run on iPad minis utilised by squadron pilots. The final part of the project is to conduct testing to quantify the improvement, if any, over the paper checklist.

4. The findings from this evaluation will be used to complete a Master of Aviation being undertaken with Massey University.

5. This project is sponsored by GPCAPT Rob Shearer, Base commander Ohakea.

Methodology

6. The research will utilise a 'between-subjects' design. The dependent variables are: workload, accuracy and speed of checklist execution, and flight path control. The independent variable is the checklist type (paper or electronic).

7. Prior to test commencement each group will undergo a short refresher on their respective checklist (paper or electronic). Each participant will then be put through two emergency scenarios in the T-6C simulator. The scenario will start at the initiation of the emergency and finish at the completion of the checklist.

8. Speed and accuracy of the checklist execution will be recorded, as well as ease of use of each checklist. Pilot workload will also be assessed.

9. The data will be analysed using the statistical program SPSS. Speed of checklist execution will be compared using a T-test, errors in accuracy will be compared using ANOVA and both ease of usability and pilot workload will also be compared using a T-test.

10. All current qualified and student T-6C pilots from 14 Squadron will be invited to participate, with up to 26 participants sought to complete the research. They will be randomly allocated to either the paper copy group or the electronic group.

Confidentiality and Ethics

11. Participation is voluntary.

12. No identifiable information will be collected from participants.

13. The NZDF Research Ethics Committee reviewed this application and endorsed it at their meeting on the 17th November 2021.

Release and Reporting

14. The results of this study will be used to complete the final part of a Master of Aviation, as well as a report to stakeholders on the findings.

15. The summary reporting will also be posted to the Human-related Evaluation and Research Repository in DDMS, so others in the NZDF can utilise the information.

16. Any substantive variations to previously approved research must have the written consent of ACOD.

Endorsement and Approval

17. This application was reviewed by DDS and no security issues were raised.

18. In my role as Chair, NZDF Research Ethics Committee, I have reviewed this project and am satisfied that all ethical and scientific requirements required by Ref A are adequately met.

19. It is therefore recommended that ACOD:

a. **Approve** the proposed research project.

J.K. HUGHES
Principal Advisor Organisational Research
DTelN Phone: +64 4 496 0141
Email: Joanne. Hughes@nzdf.mil.nz

<u>14 Squadron Checklist Trial</u>

Information Sheet

Introduction

The purpose of this study is to evaluate the T-6C Texan II checklist. The results of this study will directly contribute towards increased flight safety in the T-6C Texan II.

The research is undertaken by SQNLDR Peter Barron for contribution towards a Master of Aviation, and SQNLDR James Peters, 14 Squadron QFI.

Participant's Rights

You are under no obligation to participate. Participation is voluntary. If you decide to participate, you have the right to:

- *Decline to answer any particular question.*
- *Withdraw from the study at any time.*
- *Ask any questions about the study at any time during participation.*
- *Provide information on the understanding that your name will not be used.*
- *Be given access to a summary of the project findings when it is concluded.*

Participation in this study will not affect career progression and data collected will not be used for other means outside of this research, and for Wings course students, performance during testing will not affect, or contribute to, end of course marks.

Project Contacts

- *SQNLDR Peter Barron*
- *SQNLDR James Peters*

You are welcome to contact the above personnel if you have any questions about the project.

NZDF Organisational Research Ethical Approval

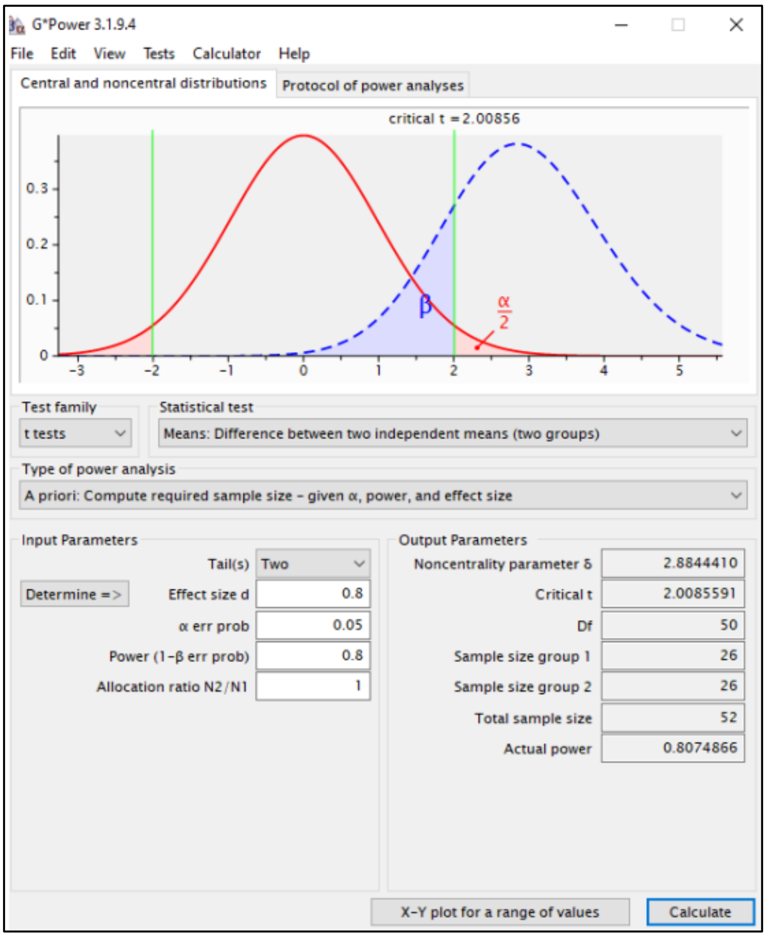
This project has been reviewed and approved by New Zealand Defence Force Organisational Research. If you have any concerns about the conduct of the research, please contact OrgResearch@nzdf.mil.nz.

I consent to participating in this study.

*Signed:*________________________________

*Date:*__________________________________

Appendix E: G*Power Analysis



Appendix F: Qualitative Data Summary

General observations

- All EEC participants pulled and reset the circuit breaker or discussed the reasons that they wouldn't reset. Many QRH participants missed the fact that the circuit breaker even needed to be reset.
- Content is a factor with both checklist types
- Head down time was a factor in both checklist types.

Checklist Content Issues (common to both checklist types)

Poor/confusing wording

(20 occurrences) -some in the list below occurred several times

- Examples
    - Went to emergency gear extension checklist due poor checklist wording.
    - (CBs – pull out/reset)
    - Conditional statements in the checklist are very poorly worded and hard to understand.
    - Poor conditional statement – If IAC/MFD failure persists…
    - Misread step 4. Confusing wording. Didn't pull and reset CB even though he knew at some stage the checklist would ask to pull CB. Even misread in the sim while stable.
    - "IAC failure persists" - poor wording. Misread as IAC failure does not persist, which meant CB not pulled.
    - Item omission – not pulling CB – misread checklist, poor wording. Interpretation of checklist.
    - Landing with unsafe gear indications confusing wording.
    - Did not conduct entire checklist due Note 6N re unable to reset CB. Participants rationale was that it was not popped, therefore unable to reset. Ambiguous content.
    - Long time spent dwelling on steps 5-6 around clarifying wording of circuit breaker.
    - Have to really read it to try and understand what it is telling you.
    - There is a US Navy (USN) term in the checklist that is not understood (RDO = Runway Duty Officer). Having a USN abbreviation in an international checklist is poor.

Bulk

(6 occurrences)

- Sheer amount of notes and content
- Not reading item due to too many words (…"safe gear down indications are…").
- Superfluous language in checklist
- Checklist is verbose.
- Landing gear malfunction is a really long procedure
- Checklist wordy, have to reread sometimes to understand what it's trying to tell you.

Physical Properties and Usability

QRH	EEC
• Checklist is a loose item which took two hands to control whilst under emergency conditions. • Almost fell off knee. • Unsafe to use – FOD • Attention is required to keep checklist open on leg. • Book/Unattended/FOD - not good to use • Loose item under G/manoeuvre • Bulky book, have to flip and then find where you are again. • QRH hard to manipulate in terms of physical size and the note location • QRH manipulation is fairly easy under 1 g and no bumps in the sim. In aircraft a different story, especially during side slips, g, etc. • Takes so much time to manipulate the checklist that it makes the scenario so much longer. • Checklist fell off knee once • Checklist almost fell off knee once • Hard to manipulate the switches having to hold on to checklist. Have to switch hands. • Sat checklist on RHS under arm to hold down on kneeboard • Flying left handed a lot which was a compensation for holding a loose item (QRH). • Adjusted flying behaviour to compensate for holding checklist (flying with left hand). • Had to cross hands to get out of side pocket. • Left handed, spent more time flying with left hand than normally would with IEC. • Hands of stick time to physically manipulate checklist x 2 in IMC • Took hand off stick to open checklist (in IMC). • Checklist stowed in canopy when hand needed during checklist management. • Had to put checklist away and get it out • Had to put checklist away and get it back out during the scenario to conduct manoeuvring • Hands off stick time to physically manipulate the checklist. • Conducting some items from memory (for manoeuvring items). • Flipping between pages/notes. • Having QRH on knee and getting the light to shine on it is hard. Hard to see in lighting conditions. • Cannot see entire checklist unless it is folded out so both pages are visible (but then not a useable size and hard to hold), otherwise have to flip over.	• When queried if double tap is annoying with gloves, participant stated that they were old gloves, and that other apps used in flight require double taps. • Clunky double tap

Checklist Errors

Error type	QRH	EEC
Loss of place	17	0
Item conducted out of sequence	8	0
Checklist manipulation issues	28	2 (double tap with gloves). Clunky double tap

QRH Issues

Miscellaneous

- Hard to know where a particular checklist starts due the way in which it's worded.

General Observations/Comments

- It is hard to find the correct checklist.
- Used finger to keep place
- Finger to keep place in checklist.
- Experienced pilots troubleshoot first before initiating checklist.
- QRH users read notes before they complete the step - trained. And also shows relevant notes at the end of the step vs EEC where notes are listed after the step, so no association.
- With QRH, participant knew to read item, then read W/C/N before actioning item.
- Participants observed paraphrasing notes. Combination of familiarity and verboseness of checklist.
- Blank page on page EA-6 was confusing.
- The steps on the checklist are hard to define (scenario 1)
- Hard to get any flow going with QRH. Physically trying to read impacts the flow.
- Holding checklist up to reduce heads down time.
- QRH CAS messages are listed right to left in the back of the book. Not intuitive. Could not find IAC 1 FAIL CAS in the QRH. Had to use IAC 2 FAIL reference.
- Had to really look for circuit breakers and/or guess where they are.
- Tempting to not use it and go straight to what you think the checklist will eventually tell you (i.e., just ignore checklist and go straight to reset the circuit breaker).
- Would not pull QRH out for minor emergencies at all.
- Reluctance to read notes
- Concerned about length of time to complete emergency.

<u>EEC issues</u>

Miscellaneous

- Initiated wrong checklist as progressed through the checklist (landing with unsafe gear indications)

Overall EEC HF Errors

- Not completing checklist item (8 occurrences)
- Single click to complete item (forgetting to double tap)
- Pushing choice box instead of decision points.
- W/C/N before action is confusing (3 occurrences)
- Formatting
    - Went to landing gear emergency extension early due note running into separate line "…if unable to lower gear… (-new line-) …handle.
    - gear/handle on two lines.
    - Notes after checklist item (e.g., lamp test, and then note about checking gear lights illuminate)
    - Confusion due to CB wording ("check in/reset"). Then, note explaining after the item (Item 5).
    - W/C/N before action is confusing
- Circuit breaker diagram
    - Aft CB panel was looked for (CB diagram needs to be clearer, or fixed through training)
    - CB diagram - collared CBs catch the eye rather than the targeted CBs
    - In CB diagram, collared CBs draw eyes to incorrect CBs potentially (for scenario 2).

Participant EEC suggestions

- Is there any value in being able to open and close notes on the EEC?
- Should be able to activate choices without having to green out everything beforehand.
- CAS menu does not need descriptors. Should just have CAS listed in two columns for easy finding. People do not read the descriptor anyway
- EEC checklist arrow leads user to click it (the links look like a checklist step and in the absence of other checklist steps on the screen below, it could lead the user to go to the link and use the wrong checklist.
- Auto scroll to put completed item at the top of the page (function trialled previously but not desirable to the amount of content that is scrolled past to get to the next item).

EEC Observations

- Participant collapsed notes once read.
- Participant collapsed notes.
- Did not use collapse notes function
- Scrolled ahead in checklist to see checklist progression. Scroll around to check pathway.
- Missed IAC malfunction in menu, found it in CAS page
- Couldn't find checklist through contents page. Had to go through CAS Index.
- Searching for IAC FAIL checklist using CAS page. Didn't realise in alphabetical order – can be remedied during formal training.
- Liked checklist place holders (magenta box)

- Clunky double tap (sometimes it works, sometimes it doesn't). Should allow for fast double tap.
- Likes choice function
- Clicking isn't intuitive for airmanship actions (e.g., below 150kts)
- Liked choice function
- Unused to format.
- Likes the choice function
- Some participants wore gloves
- Participants are more familiar with checklist on second scenario.
- Some participants read notes before conducting item, others conduct item then read notes.
- EEC makes people work according to the direction in the checklist. For example, people are testing lamp switch, and then reading note which describes why to do it, and then are going back to do it again. With the QRH, people read the notes before conducting the action as the notes are associated with the step instead of separate as they are in the EEC.

Digital vs paper-based checklists in high performance single pilot aircraft : a mixed methods investigation : a 190.895 (60 credit) research report presented in partial fulfilment of the requirements for the degree of Master of Aviation at Massey University, Palmerston North, New Zealand

Barron, Peter

2022